DOCTOR OF PHILOSOPHY

Gaze cues and language in communication

MacDonald, R. G.

Award date:
2014

Gaze cues and language in communication

R. G. Macdonald, MSc

This thesis is submitted in part fulfillment of the requirements for the Doctorate of Philosophy in Psychology

University of Dundee, October 2014

Contents

List of Figures

Acknowledgements

During the three-years I've worked towards this thesis (and indeed the seven-years I've studied Psychology at Dundee) I've been lucky to receive the help and guidance of some very decent people, whom I'm obliged by convention to make note of in the present section.

Firstly, of course, I should thank Ben Tatler. Ben worked hard to get me a position at Dundee and has always been very supportive of my (often half-baked) ideas. His brand of supervision has been ideal for me and played no small part in how much I've enjoyed the last three years.

Most of the academic, admin and technical staff at Dundee have provided me with some form of guidance throughout my time at the University, for which I'm very thankful, but I should give special mention to past supervisors Alissa Melinger and Wayne Murray and my second supervisor Nick Hopkins, who have all been enthusiastically helpful whenever I've knocked on their doors.

Big thanks must also go to past and present members (and associates) of the Active Vision Lab for years of engagement with (and constructive criticism of) my work: Yoriko Hirose, Sara Spotorno, Matt Stainer, Sharon Scrafton, Steven Savage, Clare Kirtley, Katy Inglis, Anne Scrymgeour, Riina Pievilainen and Nick Wade. Also, I must mention the enthusiastic undergraduate volunteers who helped me collect and code data: Nicole Spittle, Julia McVean, Anne MacGregor, Gemma Mackintosh, Kevin Gachee and Ahmet Zeybek.

I must also mention other friends and colleagues who have made working in the School of Psychology such fun: Katja Suckow, Gavin Revie, Laura Wakeford, Adele Maxwell, Glenn Williams, Kirsty Miller, Tibor Farkas, Samantha Swartzman, Chris Coyne, Anna Dobai, Jaime Benjamin, Jana Speth, Clemens Frenzel, Juliet Wakefield, Shane Lindsay and Anue and Mike Baker-Kukona.

Naturally, my parents and family deserve endless gratitude for dealing with me for such a very, very long time.

Finally, I need to give the most profound thanks to my partner Laura, who has been fantastically loving, supportive, patient and understanding for as long as we've been together.

Declaration

I, the candidate, declare that I am the sole author of this thesis and have consulted all cited references personally. I carried out the work of which this thesis is a record and have not previously submitted this work for a higher degree.

Ross Macdonald

Signature   Date

Abstract

During collaboration, people communicate using verbal and non-verbal cues, including gaze cues. Spoken language is usually the primary medium of communication in these interactions, yet despite this co-occurrence of speech and gaze cueing, most experiments have used paradigms without language. Furthermore, previous research has shown that myriad social factors influence behaviour during interactions, yet most studies investigating responses to gaze have been conducted in a lab, far removed from any natural interaction. It was the aim of this thesis to investigate the relationship between language and gaze cue utilisation in natural collaborations. For this reason, the initial study was largely observational, allowing for spontaneous natural language and gaze. Participants were found to rarely look at their partners, but to do so strategically, with listeners looking more at speakers when the latter were of higher social status. Eye movement behaviour also varied with the type of language used in instructions, so in a second study, a more controlled (but still real-world) paradigm was used to investigate the effect of language type on gaze utilisation. Participants used gaze cues flexibly, by seeking and following gaze more when the cues were accompanied by distinct featural verbal information compared to overlapping spatial verbal information. The remaining three studies built on these findings to investigate the relationship between language and gaze using a much more controlled paradigm. Gaze and language cues were reduced to equivalent artificial stimuli and the reliability of each cue was manipulated. Even in this artificial paradigm, language was preferred when cues were equally reliable, supporting the

idea that gaze cues are supportive to language. Typical gaze cueing effects were still found, however the size of these effects was modulated by gaze cue reliability. Combined, the studies in this thesis show that although gaze cues may automatically and quickly affect attention, their use in natural communication is mediated by the form and content of concurrent spoken language.

Chapter 1:

Introduction

When collaborating with another we primarily use spoken language to communicate. However, this is often facilitated by non-verbal cues. Arguably the most well researched of these cues is the cue provided by the gaze of another. These gaze cues are unique in that unlike other non-verbal cues they can directly inform a listener as to the focus of a speaker's attention.

Researchers across a wide range of disciplines have investigated gaze cues, including cognitive, social and developmental psychologists, evolutionary biologists, psycholinguists and computer scientists. A common finding amongst researchers is that humans have a tendency to follow the gaze cues of another (Emery, 2000). Some evolutionary psychologists have theorised that the unique contrast visible in the human eye has evolved so that we can easily detect where others are looking (Tomasello, Hare, Lehmann & Call, 2007). Behavioural experiments have shown that people are highly skilled at detecting eye movements of others and are sensitive to small changes in gaze directions (Anderson, Risko & Kingstone, 2011).

There has long been evidence that our ability to follow the gaze of another arises in infancy (Scaife & Bruner, 1975). In infants, this behaviour has been interpreted as an example of joint attention between infant and adult. These findings challenged Piaget's (1972) concept of the egocentric child and would later be important to informing key concepts and theories in child development, such as Theory of Mind (Baron-Cohen, 1995).

The present thesis investigates the relationship between spoken language and gaze cues in natural social interactions. As research on gaze cue utilisation is widespread across several disciplines and has been used to address numerous diverse questions, it is beyond the scope of this review chapter to provide a comprehensive overview of this literature. Instead, this chapter focuses on research findings necessary to inform a thorough investigation of the interaction between language and gaze cues in natural interactions. Specifically, discussion focuses on gaze behaviour in real-world interactions, the attentional effects of gaze cue stimuli and, crucially, previous work investigating gaze cues and language together. Reviewing this literature together lays the groundwork for the exploration of the relationships between gaze cues and language presented in this thesis.

Early work on gaze and social interaction

Throughout the nineteen-sixties and –seventies, before reliable portable eye-tracking systems were developed, social psychologists and anthropologists used a variety of methods to investigate how gaze was used in natural social interactions (Argyle & Cook, 1976). These studies used coarse measures relative to those recorded with eye-trackers and so could not be used to measure the precise timings of eye contact or gaze following events accurately. However, the findings of these social gaze studies can still provide insight into gaze behaviour during natural communication. The work detailed below explores the things people can infer from the gaze behaviour of another, as well as a spectrum of social factors that modulate the eye movement behaviour of individuals in a social interaction. Collectively, these studies highlight a number of the fundamental principles of gaze behaviour during social interactions and these principles have been used to help inform and interpret the studies in the present thesis.

When viewing the faces of others, people are able to assess emotional state as well as infer personality traits. Tomkins and McCarter (1964) presented participants with images of faces simulating emotional expressions. Participants identified the majority of these correctly, and those that were incorrectly identified were commonly misidentified across participants, showing a shared and astute ability of participants to identify the emotional state of others. Kleck and Nuessle (1968) asked participants to make judgments about people interacting in custom-made films in which actors looked at each other (mutual gaze) either 15% or 80% of the time. When there was low mutual gaze, participants used negative attributes to describe the actors, such as "cold", "defensive" or "submissive". High mutual gaze resulted in the use of much more positive words, such as "friendly", "natural" and "sincere". A similar study, using a real interactor rather than a filmed duo, investigated this effect quantitatively (Mehrabian, 1968). Participants viewing an interactor rated how much they thought he liked the person with whom he was interacting. Likeability scores were found to increase with the amount of eye contact provided in the interaction. These early studies show that when viewing others, people use the facial expressions and gaze behaviour of those others to make inferences about their personality and mood.

Early research investigating gaze behaviour during real interactions found a negative correlation between proximity and the amount of eye contact during a conversation (Argyle & Dean, 1965). In this study a confederate and participant sat at 90° to each other around a square table and discussed images on a Thematic Apperception Test (TAT) card. The confederate gazed at the participant's eyes throughout these conversations and between conversations the proximity of the confederate to the participant was varied. Two observers behind a one-way screen used stopwatches to time the participants' glances toward the confederate, providing a measure of mutual gaze. The experimenters found that the closer the confederate, the shorter the total eye contact time between the interactors. Argyle and Dean argue that this is due to the role eye contact and proximity play in intimacy. The combination of close proximity to another and a high level of mutual gaze is characteristic of an intimate social interaction, so in this experimental interaction with a stranger, participants avoid mutual gaze. However, when the confederate is less proximal mutual gaze is less intimate, therefore the participants do not need to suppress glances to the confederate. In the same paper, Argyle and Dean report the findings of a "perceptual experiment": Participants were brought into a room with an inanimate object on display and instructed to stand "as close as is comfortable to see well". After this they were presented with three different displays, one after the other: 1) a cardboard cutout of the lead experimenter's face, looking at the participant, 2) the lead experimenter, looking at the participant and 3) the lead experimenter with his eyes closed. For each display participants positioned themselves as they did for the initial display. The authors found that there was no significant difference in the distance participants positioned themselves from the closed-eyes researcher and the cardboard cutout. However, they were found to stand significantly farther away from the researcher when he was looking them in the eye. These studies show that our proximity to another person affects the extent to which we engage in mutual gaze with that person, suggesting that our gaze orienting behaviour can be modulated by the context of the social interaction in which we are involved.

As well as the proximity to another, the amount of mutual gaze in an interaction can be influenced by the non-verbal signals and social perceptions of another. Efran (1968) conducted a study in which college Freshman participants talked with two confederates, who were either both Seniors, both Freshman or one of each. One confederate always smiled more and was more approving of the participant's comments than the other. The participants were found to look more at the approving confederate overall, suggesting that people engage in more eye contact with those people they feel positive towards, supporting earlier findings from similar studies (Efran & Broughton, 1966; Mehrabian, 1968). Other studies have also shown that just as people look more at those giving positive signals, they will decrease the extent of mutual gaze when involved in negative interactions, such as those in which difficult topics are discussed (Exline & Winters, 1966) or those in which the other interactor is highly critical (Modigliani, 1971).

The level of approval of the confederate was not the only factor to effect mutual gaze in Efran's (1968) study; increased mutual gaze with the approving confederate was not found across all confederate combinations. When the approving confederate or both confederates were Seniors there was a significant effect of the approving signals, however when the approving confederate or both confederates were Freshmen, this effect was not found to be significant. These results suggest that whether the approval of another increases mutual gaze depends on how that person is perceived socially; mutual gaze is increased when the approving individual is perceived as having a higher social status (a college Senior) compared to equal social status (a Freshman). In a review of early work on social status and gaze, Ellyson, Davidio and Fehr (1981) reported that the relationship between social status and gaze is more complicated. They discussed work by Exline, Ellyson and Long (1975) that found interactors will typically look more at another while listening compared to when speaking. Ellyson et al (1981) discuss this in light of their research showing that dominant interactors (those with a higher social status or expert in a discussion) spend approximately equal amounts of time looking when speaking and listening. Therefore, higher social status is associated with a higher look-while-talking to look-while-listening ratio. This may be because the dominant individual looks more when speaking to make sure that the subordinate is listening, while the subordinate looks more when listening to ensure that they understand the dominant individual or to signal to the dominant individual that they are listening.

Cultural differences in gaze behaviour were also explored during this period. Watson and Graves (1966) investigated differences in behaviour during social interactions between Arab and American students in Colorado. Pairs of either Arab or American friends were observed behind a one-way mirror as they conversed for 5 minutes. Raters behind the mirror scored each participant minute-by-minute on a number of non-verbal metrics, including gaze behaviour measured on a 1-4 scale, with 1 denoting direct eye contact and 4 denoting no visual contact. Arab students were found to have significantly lower scores than the American students, due to increased levels of mutual gaze. Similar work investigated differences in social gaze between racial sub-groups within the same nation; specifically, mutual gaze differences between black and white Americans (LaFrance & Mayo, 1976). In a frame-by-frame analysis of a black-black and a black-white conversation, LaFrance and Mayo reported that the white participant looked significantly more during listening than the black participants. In a second study, multiple dyads (half black, half white) were covertly observed in a natural setting by observers who timed mutual gaze events with stopwatches. Black pairs were found to spend less time looking when listening than the white pairs. The results of these studies should, however, be interpreted cautiously as Watson and Graves (1966) gained a small amount of data from a small sample of participants and these data were generated from a fairly subjective rating system. Furthermore, there is no assurance in the paper that raters were blind to the hypothesis. LaFrance and Mayo (1976) only used three individuals in their frame-by-frame analysis and their method of timing mutual gaze clearly lacks precision.

The above experiments focussed on eye contact between individuals, rather than gaze cueing or following. Due to the technical limitations of the time, controlling and measuring eye contact was a lot easier than attempting to investigate gaze cueing utilisation precisely and accurately. However, Milgram, Bickman and Berkowitz (1969) were able to investigate gaze following in a large-scale observational study. Milgram et al. used groups of one, two, three, five, ten or fifteen confederates who were instructed to look at a specific 6th floor window from a spot on a busy New York City sidewalk. The researchers analysed film recordings of these events and calculated the percentage of pedestrians that followed the gaze cues of the group. Approximately 40% of pedestrians looked up when one confederate was looking at the window, and this percentage increased as group size increased up to five confederates and then plateaued at around 80% for larger groups. These results could be explained in terms of low-level orienting of attention; as group size increased, more gaze cue stimuli were present, meaning more people were likely to attend to these cues and follow them. This explanation would predict that significantly more participants would follow the gaze cues of a crowd of fifteen compared to a crowd of five, however this was not the case. An alternative explanation would be that participants make a top-down assessment of whether or not to look where others are looking depending on the perceived usefulness or interest of doing so. It is possible that five individuals looking up at one spot is a rare enough occurrence for participants to conclude that it is worthwhile to follow their gaze. The addition of extra confederates may then have little, if any, additional effect on the participants' decisions. Whether determined by the low level stimuli or the top down decisions of the participants, Milgram et al. (1969) showed that as the number of gaze cues directed to a location increases the likelihood that the location will be fixated increases.

The studies detailed in this section show a high degree of flexibility in gaze behaviour in human social interactions. People view high levels of mutual gaze positively (Kleck & Nuessle, 1968; Mehrabian, 1968) and the extent to which people in an interaction engage in mutual gaze increases as the perceived mood of their partner increases (Efran, 1968). Conversely, mutual gaze decreases in negative interactions (Exline & Winters, 1966; Modigliani, 1971). Physical proximity has been found to negatively correlate with mutual gaze (Argyle & Dean, 1965) and the perceptions of the social status of the person with whom one is interacting have complex effects on how one looks at them (Ellyson et al., 1981). Together, these findings show mutual gaze is modulated by the interactors' physical position, the behaviour of their partner and their beliefs about their partner. These findings raise questions about how these factors affect communication. Argyle described gaze as "both signal and channel" (1988, pp.153) because gaze behaviour is unique amongst non-verbal cues, as the organs involved in providing cues are also required to perceive cues. Therefore, the more one engages in mutual gaze with their partner, the more opportunity one has to perceive their partner's gaze cues.

The findings of these early studies provide a foundation for research on social gaze using more precise and accurate measures, including experiments focussing on directional gaze cues. Although the mostly naturalistic social experiments detailed above show evidence for selective gaze utilisation in human interaction, many cognitive and developmental studies using more controlled lab-based paradigms, including those using eye trackers, have argued that there is an automatic and innate component to the utilisation of gaze cues in social interactions.

The gaze cueing paradigm

The approach to researching gaze cueing in the cognitive science domain has varied drastically from the approach of the early social psychologists. The cognitive approach has been to use artificial gaze cueing stimuli in highly controlled lab based environments to identify small, but significant attentional effects of gaze cues using behavioural and eye-movement measures. Much of this research has used the framework of Friesen and Kingstone's (1998) gaze-cueing paradigm and has provided a great deal of insight into how we respond to gaze cues, as well as the factors that influence these responses.

Friesen and Kingstone (1998) based their gaze-cueing paradigm on the Posner cueing task (1980) and used it to investigate the effect of gaze cues on attention. The paradigm involved a target (a letter –"F" or "T") on the left or right of a display screen. In the centre of the screen there was a simple drawing of a face looking left, right or centrally, which participants were informed did not predict target location. In this experiment, participants were required to detect, localise or identify the target as quickly as possible. Reaction time differences between gaze cue conditions were used to infer the extent of attention capture. Participants were quicker in all three tasks when the cue was directed to the valid location compared to when it was neutral or to an invalid location. The cueing effect was found at short stimulus onset asynchronies

(SOAs) (105ms), but disappeared at long SOAs (1005ms). This time course, combined with the effect occurring despite participants being told to ignore the gaze cues, provides strong evidence that participants reflexively followed gaze cue stimuli. This suggests that the brain is specialised to cause an attentional shift in the direction of a gaze cue. One can speculate in a number of ways as to how this mechanism may have arisen, however the prevailing theory is that, like the unique phenotype of the human eye (Tomasello et al., 2007), the tendency to follow the gaze of another evolved to aid communication and collaboration (Emery, 2000).

In the same year as the Friesen and Kingstone (1998) paper, Hood, Willen and Driver (1998) published striking results using a version of the gaze cueing paradigm with 10-28 week old infants. The experimenters found that that these infants were quicker to look in the direction of a visual probe when they were provided with an artificial gaze cue in the congruent direction compared to a cue in the incongruent direction. This evidence for a gaze cue effect in children as young as three-months provides strong support for an innate and automatic attentional shift in humans caused by directional gaze cues.

One obvious alternative explanation for the results found in the experiments above is that participants were responding to simple directional cues embedded in artificial gaze stimuli rather than actually responding to the social gaze cue. It may have been incidental to performance that these directional cues took the form of a gaze cue; it may simply have been that people respond in this way to low-level visual stimuli. To explore this issue, Ricciardelli, Bricolo, Aglioti and Chelazzi (2002) carried out a series of eye-tracking experiments using the gaze-cueing paradigm with either photographs of gaze cueing faces or simple arrow stimuli. Saccade error measures for gaze cue stimuli were in line with typical reaction time results for gaze cue paradigm studies; there were more saccade errors when incongruent stimuli were used compared to congruent stimuli, but only at short SOAs. However, this effect was not found for arrow stimuli, providing support for the hypothesis that people automatically orient attention to gazed-at locations.

The results of Ricciardelli et al. (2002) were not found in all studies investigating the differences between gaze cues and other symbolic cues. Tipples (2002) found that the distracting effects of arrows were no different from the effects of gaze cues. One reason for the discrepancy between the findings in these two studies may be the arrow stimuli used. Ricciardelli et al. (2002) created arrows by erasing half of the central “X” fixation cross in the centre of the screen, whereas Tipples (2002) used arguably far clearer and more typically encountered arrow stimuli. It may be that the lack of an attentional orienting effect from arrows in the former study was due to participants taking more time to identify the stimuli as arrows. A later study (Kuhn & Benson, 2007) attempted to replicate the experiments of Ricciardelli et al., again with less subtle arrow cues, and found no difference between the effect of arrow and gaze stimuli. However, when focussing on the trials that resulted in saccades in the incorrect direction, the experimenters found that these saccade latencies were significantly quicker than those in the congruent direction for gaze cues trials, but not arrow trials. Kuhn and Benson (2007) argued that although arrows and gaze cues resulted in the same number of erroneous saccades, the erroneous saccades caused by social-biological gaze cues were more reflexive in nature than the erroneous saccades caused by social-non-biological arrow cues. This idea that gaze cues elicit a more reflexive response than arrow cues is in line with the results of earlier work showing that when cueing highly unlikely target locations, gaze cues shifted attention reflexively, whereas arrow cues did not (Friesen, Ristic & Kingstone, 2004; Langdon & Smith, 2005).

Although there are a number of similarities in the behavioural responses to gaze and arrow cues, the above experiments highlight some key differences that suggest the use of unique brain systems for gaze following. Further evidence for these unique systems is found in a pair of papers investigating gaze cueing in split-brain patients. Kingstone, Friesen and Gazzaniga (2000) found a slower response time in trials with incongruent gaze cues in the left-visual field only, indicating that the reflexive gaze following is localised in the right hemisphere. In a later paper, split-brain patients carried out the same experiment but with arrows (Ristic, Friesen & Kingstone, 2002). Incongruent arrows cues disrupted response time in both the left and right visual fields, suggesting that these attentional shifts use different neural systems than those caused by gaze cues.

The neural systems involved in gaze following have been shown to have top down influence on how people respond to gaze cues. Ristic and Kingstone (2005) conducted a Posner-type gaze cueing task using ambiguous central visual stimuli. These stimuli were simple images that could be identified as a strange-looking car or a person in a funny hat. A key feature of these stimuli was that they featured a pair of white-filled and black-bordered circles, each with a black-filled circle inside (Figure 1.1). The smaller black circle was either positioned to the left side or right side of the larger white circle. These circles either represented the wheels of the car or the eyes of the person in the hat. The experimenters either told the participants that the stimuli represented a car or a person.

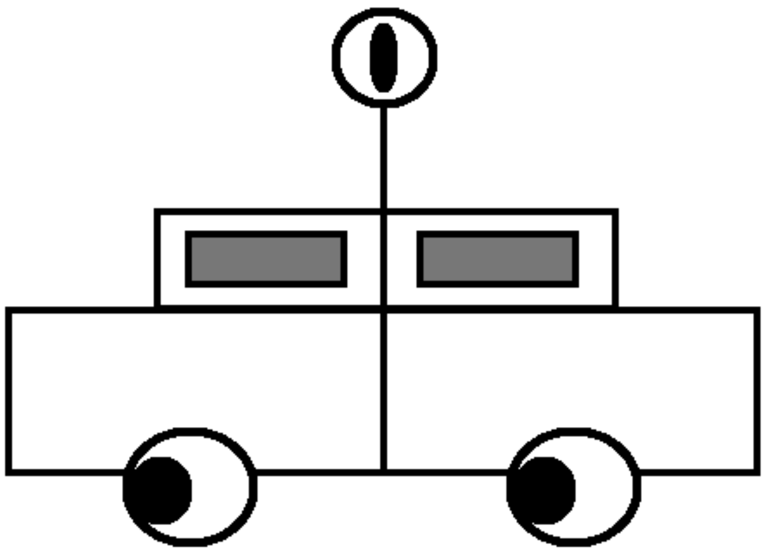

Figure 1.1. A reproduction of the ambiguous central visual stimulus used by Ristic and Kingstone (2005). Participants were informed that the stimulus was either a strange-looking car or a person in a funny hat. The actual stimulus used can be seen in the original paper.

In a third between-subjects condition, participants were provided with schematic faces looking to either the left or right instead of the ambiguous stimuli. In all cases participants were instructed to ignore the stimuli and to complete the task as quickly and accurately as possible. Participants in the schematic face and person-in-hat condition displayed the standard gaze cueing paradigm behaviours; they were significantly slower to respond when the gaze cues were incongruent to target location. However, despite having precisely the same visual characteristics as the stimuli in the person-in-hat condition, there was no cost of incongruent wheel cues to target location in the strange-looking-car condition. This evidence strongly suggests that the distracting effect that gaze cues have on attention is not simply a bottom-up process driven by the physical characteristics of gaze cue stimuli, but rather requires a top-down understanding that the stimulus being perceived represents the gaze of another.

It is also possible that the key factor in the distracting effect of the person-in-hat stimulus is not that the eye position represents a gaze cue or even a biological cue, but that the eye position represents a general spatial cue. The location of the black circle within the wheels of the car does not represent a cue, as the position of the wheel in its axis has no relation to the direction of the car. Regardless of whether the distracting effect of the incongruent person-in-hat stimulus is caused by a distracting social cue or a more basic representational spatial cue, it is clear from these results that the visual features of a gaze cue stimulus are not sufficient to cause a distracting gaze following effect. For this effect to occur, the viewer must recognise the stimulus as a cue.

The potential effects of visual properties of gaze stimuli and higher-level cognitive factors have lead some researchers to postulate or explore the possibility that there are two or more attentional streams involved in the response to a gaze cue. Rather than exploring the difference between gaze cues and social-non-biological arrow cues, Downing, Dodds and Bray (2004) inventively compared gaze cue stimuli to a non-social-biological cue: the human tongue. Participants carried out a typical gaze cueing paradigm, either with gaze cues or a tongue pointing left or right. Much like the comparisons with arrows, these tongue stimuli had broadly the same effect on attention as gaze cue stimuli. One difference found however, was that at very low reliabilities tongue stimuli did not distract participants, but gaze cues at the same reliabilities continued to capture attention. This provides evidence that the unique characteristics of a gaze cue may not be their automatic cueing effect, but rather the gaze cues' resistance to top-down influence. Specifically, the persistence of the attention orienting effect of gaze cues at low reliabilities. It may be the case that the reflexive response to gaze cue stimuli may be a general response to spatial cues, common to gaze cues, arrows, tongues and other cues, whereas a top down selective attentional stream may be unique to gaze cues.

A number of studies have altered the reliabilities of gaze cue stimuli to find evidence for two attentional processes involved in gaze following. Driver et al. (1999) used a gaze cueing task similar to that of Friesen and Kingstone (1998), but reduced the likelihood of the gaze cue being congruent to 20%, therefore making the best strategy in the task to attend to the side not cued by the gaze stimulus. Driver et al. (1999) found that even in this situation, incongruent gaze cues slowed reaction time at SOAs of 100 ms and 300 ms. However, at an SOA of 700 ms this difference was reversed; response time was significantly slower in the congruent condition compared to the incongruent condition. A later eye-tracking study (Kuhn & Kingstone, 2009) supported this finding: with a gaze cue stimuli reliability of 20%, incongruent trials resulted in more saccade errors at short SOAs (0-100 ms), but the reverse was true at long SOAs (900 ms). In both studies, there was a benefit to congruent trials at all SOAs when gaze cue reliability was at 50%. Using a different paradigm, Hill et al. (2010) found that response time in a task with 0% gaze cue accuracy only differed from performance in a task with 100% gaze cue accuracy at short SOAs, whereas congruent trials were responded to faster than incongruent trials at all SOAs in a task in which gaze cue reliability was 50%. All of these studies provide evidence for reflexive and flexible attentional mechanisms involved in gaze following. The task disruption at low SOAs when cues are highly unreliable suggests a short-lived and automatic attentional shift. The evidence for an adaptive gaze cueing effect at longer SOAs for unreliable cues suggests that slower, top-down inhibition occurs when we perceive gaze cues to be unreliable.

The evidence for top-down influence on gaze following has large implications for the way in which gaze cueing should be investigated. There is good evidence that gaze cue stimuli out of context (Ristic & Kingstone, 2005) or with low perceived reliability (Hill et al., 2010) can decrease the extent to which we follow these cues. Given that in the real world we can be provided with gaze cues in a wide range of contexts and with varying reliabilities, it may be that understanding these top-down effects is more important to understanding how we use gaze cues than the low level, short-lived attentional shift that is the focus of many lab-based studies. However, some experiments using controlled gaze cue paradigms have provided insight into these higher-level effects on gaze following.

There is evidence that participants infer the perspective of the person represented by gaze cueing stimuli and that this affects gaze following. Previous research has shown that turning gaze-cueing face stimuli upside down supresses the typical gaze cueing effect (Langton & Bruce, 1999), however when Bayliss and Tipper (2006) rotated the face stimuli to 90° (so that left and right gaze cues were spatially cueing up and down) participants were found to have longer response times when the target was in the opposite direction to where the face would be cueing, were it upright. These results show that participants' attention was not only shifted to the gazed at location, but to where they inferred that the face would be looking. This suggests that the top-down interpretation of the meaning of a gaze cue influenced the attentional response to gaze cue stimuli at an early stage.

Further evidence for the influence of gaze cue context comes from Ricciardelli, Carcagno, Vallar and Bricolo (2013), who found that incongruent gaze cues lead to saccade errors only when the stimuli cued distractors that were potential targets (valid distractors). When incongruent cues were directed to either distractors that were not potential targets (i.e. distractors that were positioned in locations in which a correct target never appeared) or to nothing, these cues did not lead to saccade errors. Once again, this shows the surroundings in which a cue is given affects the attentional response to gaze cue stimuli, indicating a selective component of gaze following.

Gaze following in gaze cueing tasks can be affected by the person following the cue, as well as the person providing the cue. Bayliss, di Pellegino and Tipper (2005) found that the distracting gaze cueing effect was stronger in females than males and that effect magnitude was also negatively correlated with scores on the Autism-Spectrum Quotient Questionnaire (Baron-Cohen, Wheelwright, Stott, Bolton, & Goodyear, 2001). Participants carrying out a gaze-cueing task in Italy were shown distracter gaze cueing stimuli made from the faces of Italian political figures, including Silvio Berlusconi (Liuzza et al., 2011). The gaze of Berlusconi was found to cause significantly more interference in the task for right-wing voters (in-group) than left wing voters (out-group). These results suggest that people may be more prone to following the gaze cues of others with shared beliefs. Reciprocally, people have been shown to rate faces as more trustworthy if these faces provided congruent cues rather than incongruent cues in a gaze cueing task (Bayliss, Griffiths & Tipper, 2009). The studies outlined above show that there are differences in the attentional effects of a gaze cue depending on who is following the cue and what they think of the person providing the cue.

The experiments discussed above highlight the important and diverse insights provided by the Posner-type gaze cueing paradigm into how we interpret and respond to gaze cues. Replications and variations on the original paradigm (Friesen & Kingstone, 1998) have consistently found that gaze cues capture attention even when participants know they are uninformative. It is clear from this that there is an automatic shift of attention caused by these cues, however whether this shift is caused uniquely by gaze cues is another issue. Distracting non-biological (Tipples, 2002) and non-social cues (Downing, Dodds & Bray, 2004) have broadly the same reflexive effects on attention, suggesting that this phenomenon is common to a broad range of representational spatial cues. There is evidence that gaze cues do have unique attentional effects, but the differences between gaze cues and other cues are found not in the immediate reflexive response to the gaze cue, but in the persistence of gaze cues at low likelihoods (Downing, Dodds & Bray, 2004; Friesen, Ristic & Kingstone, 2004; Langdon & Smith, 2005). Furthermore, the wealth of evidence that gaze following in these tasks is influenced by top-down factors, such as the interpretation of the meaning of gaze cues (Ristic & Kingstone, 2005; Bayliss & Tipper, 2006) and the perception of those providing gaze cues (Liuzza et al., 2011), shows that gaze cueing is unique, but that this uniqueness is not manifest in the immediate reflexive attentional response.

One reason that the automatic attentional shift has been the focus of much gaze cueing research in cognitive science may be that the gaze-cueing paradigm effectively reduces a gaze cue to a simple visual spatial cue. Birmingham and Kingstone (2009) likened comparisons between artificial gaze cueing stimuli and arrows to “taking a

150-pound person and a 150-pound rock, weighing them, and concluding that they are the same" (p. 132). The obvious implication of this statement is that key aspects of a natural gaze cue are missing from the artificial stimuli used in the gaze cueing paradigm. The artificial cues are isolated from any communicative or social context, with which they would almost always be accompanied in the real world. Additionally, in the real world, a key aspect of gaze cue utilisation is to identify and orient to an informative gaze cue, which is not considered in the gaze cueing paradigm. The limitations of the gaze cueing paradigm have lead researchers to investigate social attention and gaze following using social scenes and real world paradigms.

Orienting to another's gaze in social scenes and the real world

During a natural gaze cueing event, the viewer must identify and orient towards the eyes of the cueing individual before following the gaze cue. The gaze cueing paradigm can inform us about what happens once a gaze cue is witnessed, but it cannot offer insight into how we orient towards the gaze cue of another. Studies using pictures of people or social scenes as well as real-world studies have been able to investigate the extent to which we focus on the eyes of others and the factors that influence our gaze orienting behaviour.

One of the earliest eye movement experiments investigating how people look at others was carried out by Buswell (1935), who showed participants, set-up with eye-tracking apparatus, a print of *'The Solemn People – Taos Indians'* by Walter Ufer. Participants were found to focus mainly on the faces of the figures in the painting, largely ignoring the landscape and bodies of the figures. Yarbus' (1967) book *Eye movements and Vision* provides a number of examples of eye movement scan paths from participants viewing a variety of isolated faces. These face stimuli include photographs and drawings of people, sculptures and animals. Common to these scan paths is a relatively large amount of time spent fixating the eyes of the face. These early eye movement studies suggest that people have a tendency to fixate the faces of others, and within these faces, people tend to focus on the eyes.

Yarbus (1967), however, provided more evidence showing that the extent to which participants looked at faces and eyes was influenced by task. Participants viewed Ilya Repin's *'An Unexpected Visitor'* under a number of different instructions. These different instructions varied the scan paths of a participant and consequently the relative time spent viewing faces. However, the scan paths presented by Yarbus show that the faces of the prominent figures in the painting were fixated in all instruction conditions. This suggests that although the task requirements change how we look at a scene, the tendency to fixate on faces and eyes is persistent. This may be because faces and eyes are often regarded as informative in a wide range of tasks.

Low-level mechanisms can arguably provided simpler explanations for why people orient to the eyes when viewing pictures of faces. Firstly, when viewing scenes, people show a central fixation bias regardless of the distribution of visually and semantically salient features (Tatler, 2007). Since the eyes are necessarily in the centre of an image of the face, a central fixation bias would result in greater total fixation time on the eyes. Bindemann, Scheepers and Burton (2009) investigated this by monitoring the eye movements of participants when viewing faces at different orientations ranging from profile to frontal view. Between 0-250 ms after the presentation of the face stimuli, fixations were focussed around the centre of the image, regardless of the orientation of the face. However, after 250 ms fixations were centred on the eye in all orientations, showing that the tendency to fixate the eyes within a face is not driven by the central fixation bias. A second low-level explanation may be able to account for these results. The eye and eyebrow region have been shown to be important in holistic face processing (Lai, Li & Wechsler, 2007), so it may be that the eyes are fixated more than other facial features in order to process or recognise faces. Levy, Foulsham and Kingstone (2013) investigated this by inventively removing eyes from the face. The experimenters created computer generated characters that were human, humanoids (non-human characters with eyes positioned on the face) or monsters (non-human characters with eyes positioned away from the face). There was no significant difference found between the character-types in terms of how much time was spend fixating the eyes. Furthermore, in all conditions the eyes were fixated early and often. Combined these experiments show that neither the central fixation bias, nor holistic face processing can fully account for the bias for eyes found when participants view images of faces, therefore this bias is likely due to the informativeness of the eye gaze of another in social situations.

The effects of task that Yarbus (1967) found on where a participant looked in a painting of people have been found in photographs of social scenes (Smilek, Birmingham, Cameron, Bischof & Kingstone, 2006). Later eye-tracking studies specifically focussed on whether the explicit (Itier, Villate & Ryan, 2007) or implicit (Birmingham, Bischof & Kingstone, 2007) informativeness of gaze cues in a task affected the extent to which participants looked at the eyes of others. In Itier et al. (2007), participants were asked to either assess the head direction or eye direction of face stimuli. Although fixations on the eyes were present in both conditions, there were significantly more initial saccades to the eyes in the eye direction condition, suggesting that participants were fixating on the eyes when they were explicitly relevant to the task at hand. Birmingham et al. (2007) presented participants with images of scenes including people. The participants were either told that they would have a memory task after viewing the scenes or they were not told anything. Participants in the group that was informed of the memory task were found to spend significantly more time fixating the eyes of the people in the scenes. This finding shows that when trying to encode a scene into memory, participants fixate on the eyes of people in the scene, suggesting that participants consider these to be highly informative areas. Together, these experiments show that the extent to which people look to the eyes of others is affected by the perceived informativeness of these cues to the task.

The uniqueness of the gaze orienting effect was investigated by Birmingham, Bischof and Kingstone (2009) in the same way that Ricciardelli et al. (2002) investigated the uniqueness of the gaze following effect: by comparing eye stimuli to arrows. Participants were shown pictures of real world scenes featuring arrows, people or both. The experimenters found that participants focused on the faces and eyes of the people in the scenes and largely ignored the arrows. Therefore, although the gaze cueing paradigm has shown arrows and gaze cues to be equivalent cueing stimuli in a number of ways (Tipples, 2002; Kuhn & Benson, 2009), the two cues differ greatly in the extent to which people spontaneously orient towards them.

Some eye tracking studies have found that social factors affect the extent to which we look at the faces and eyes of others, echoing the much earlier findings of social psychologists (e.g., Argyle & Dean, 1965; Exline & Winters, 1966). Birmingham, Bischof and Kingstone (2008) found that participants spent proportionally more time fixating the eyes of a group of three people in a static scene when they were interacting with each other compared to when they were not. This could be because viewing a social interaction elicits gaze orienting behaviour, or it could be more simply that eyes are more important to understanding a scene in which a social interaction is taking place, compared to when no interaction is taking place. Crosby, Monin and Richardson (2008) showed that participants were more likely to look at an individual on a monitor if they thought the individual could hear comments that were potentially offensive to that individual. Participants were shown a video with four front-facing men wearing headphones, separated into quadrants of the screen. Three of these discussants were white and one was black. The men in the video each made statements about their university's admissions policy, and one of the white discussants made a statement critical of affirmative action in a way that was potentially offensive to the black discussant. For half of the participants, the black discussant was wearing his headphones and for the rest of the participants, his headphones were removed (meaning he could not hear the other discussants). The experimenters found that during the potentially offensive statement participants in the headphones-on condition spent significantly more time fixating the black discussant than those in the headphones off condition. These results show that social factors such as beliefs about another individual can affect how others look at them.

All of the evidence from the lab-based experiments outlined above shows that when viewing images of people, participants look to the faces and the eyes of these people, suggesting that we are drawn to the faces and eyes of others either innately or due to the potential informativeness of the eyes. However, the stimuli used in these paradigms have an important distinction from eyes in the real world: The eyes of others in the real world can perceive the viewer. Argyle & Dean (1965) found that people will look less at a present person than a cardboard cut out of that person and more modern eye-tracking studies have also investigated the effect of social presence. Laidlaw, Foulsham, Kuhn and Kingstone (2011) found that participants spent less time looking at a present confederate than they spent looking at the same confederate viewed on a monitor. This result was suggested to be due to the participants wanting to avoid a potential social interaction with a stranger. Gallup, Hale et al. (2012) carried out an updated version of the Milgram (1969) crowd experiment (see *Early work on gaze and social interaction)*, which included an investigation of the effect of the direction in which the pedestrians approached the crowd. Gallup, Hale et al. (2012) found that participants walking towards the crowd from behind showed a higher propensity to follow the gaze cues than those approaching from in front of the crowd. In other words, pedestrians were more likely to follow the gaze cues of the crowd when the crowd could not observe their behaviour. In a similar study, Gallup, Chong and Couzin (2012) observed the gaze following behaviour of people walking past an attractive item in a hallway. They found that people were more likely to follow the gaze of somebody walking in front of them than somebody walking towards them. These two studies support the suggestion by Laidlaw et al. (2011) that people are less likely to look at and follow the gaze of an unknown individual if there is a possibility of a social interaction, or simply a possibility of being noticed. It is clear that social presence affects the way we orient to and follow gaze cues, therefore it is important to continue investigating natural interactions to understand how gaze cues are used in the real world.

The research on gaze orienting behaviour outlined above shows that people are drawn to the faces and eyes of others when viewing static scenes. This preference for eyes cannot be fully explained by low level factors such as central bias (Bindemann et al., 2009) or face recognition processes (Levy et al., 2013). Instead, the preference is likely due to the perceived informativeness of the eyes of another. This is supported by evidence that changing the informativeness (Itier et al., 2007) or social context (Birmingham et al., 2008) of static eye stimuli changes the extent to which people fixate the eyes. However, the static stimuli used in these experiments have lacked social presence, which has been shown to affect gaze utilisation (Laidlaw et al., 2011; Gallup, Hale et al., 2012; Gallup, Chong et al., 2012). Since all natural gaze cues people observe in the real world necessarily require social presence, there is a limit to what experiments using static images of social scenes can tell us about how we orient and follow gaze cues.

Real world gaze cues also almost always occur alongside spoken language. Like spoken language, gaze cues are used to communicate ideas to another. The evidence outlined in this section has shown that the informativeness of a gaze cues can effect how these cues are utilised. Perhaps the most common factor that affects the informativeness of a gaze cue is the language that accompanies the cue. Therefore to understand how gaze cues are used in the real world it is important to understand the relationship between language and gaze cueing.

Language and gaze cues

An intimate link between the language we hear and where we look has been shown by the visual world paradigm (Cooper, 1974). This paradigm involves participants listening to spoken language and viewing a scene while their eye movements are monitored. Altmann and Kamide (1999) found that not only would participants look at items in a visual scene that corresponded to the words they were currently hearing, but that participants would make anticipatory eye movements to items when they predicted these items were about to be mentioned. Using a similar paradigm, the link between gaze allocation and spoken language has been shown to extend across conversants (Richardson, Dale and Kirkham, 2007). Two isolated conversants, viewing identical visual stimuli, were found to align their gaze allocation significantly more than chance during a conversation about the stimuli represented on the screen. These results collectively show that where we look is influenced by what we hear and that this allocation of gaze can become aligned with someone with whom we converse.

As well as language influencing the allocation of gaze, the cues provided by gaze have been shown to affect language comprehension and production. Clark and Krych (2004) used an interaction between participants in a collaborative task to investigate the relationship between natural language and non-verbal cues. In their study, two builder participants made structures out of building blocks, following the instructions of a director. The builder participants were found to use gaze cues, amongst other non-verbal cues, to communicate with the director. The directors altered their utterances in response to these gaze signals, showing that gaze cues can influence the language used in an interaction.

Hanna and Brennan (2007) conducted a series of experiments specifically investigating the use of gaze cues during instructions. In the first experiment, participant pairs worked together in an object-matching task. One participant was assigned the role of director and the other the role of matcher. In each trial, matchers had an array of objects in front of them, lined up from left to right. The directors either had a mirrored array or a circular array of the same objects, as well as a non-verbal schematic, which informed them of the object to which they should direct the matcher. Matchers were required to select objects as quickly as possible following direction from the director. In experimental trials, there was always a distractor object that was very similar to the target object. Monitoring the eye movements of the matcher, Hanna and Brennan (2007) found that the matcher would use the director's gaze cues in order to fixate the target object before the point of verbal disambiguation. This effect was found to be strongest when the item arrays were mirrored. A second experiment repeated this procedure using either mirrored or reversed display (so that, from the matcher's perspective, any item on the left of the matcher's array would be on the right of the director's display). Although a larger effect was found in the mirrored condition, matchers were found to use the gaze cues of directors to identify the target before the point of linguistic disambiguation, even when directors were using the reversed display. This is good evidence that people use gaze cues to facilitate comprehension, but also that people consider the meaning of these cues from the speaker's perspective, rather than simply following a visual cue.

In addition to gaze cues facilitating communication by solving ambiguities, there is also evidence that the gaze of another can be used to infer the definition of unknown verbs. Nappa, Wessel, McEldoon, Gleitman and Trueswell (2009) demonstrated this by showing three- to five-year-olds drawings of events with two potential perspectives (e.g., a rabbit fleeing from an elephant/an elephant chasing a rabbit). A video presented above these images showed a speaker providing a gaze cue to one of the subjects in the drawing. The speaker described the scene depicted by the drawings using a non-word verb (e.g., "blicking") in a referentially informative ("the rabbit is blicking the elephant") or uninformative sentence ("he is blicking him"). The children were then asked to define the non-word. The gaze cues of the speaker were followed in both conditions and in the uninformative condition the children defined the non-word verb as the action being carried out by the gazed-at character. These findings present further evidence that people use the gaze allocation of another to infer the intentions of that other during communication.

A similar set of experiments by Staudte and Crocker (2011) used an adult population and showed participants videos of a robot describing the featural and spatial relations between a series of visible objects, while providing gaze cues. In one experiment, the robot made mistakes in its descriptions that could have been corrected in two different ways. The experimenters found that participants would correct in the way that was congruent with the gaze cue, suggesting that they were inferring meaning from the robot's gaze and assuming that the robot meant to refer to the object to which it allocated its gaze.

Another, arguably more parsimonious, explanation of the results of Nappa et al (2007) and Staudte and Crocker (2011) is that the bias for the gazed-at items is not due to the participants making inferences about the speaker's intentions, but rather that the bias is due to gaze cues orienting attention to the gazed-at location (Staudte, Crocker, Heloir, & Kipp, 2014). This attentional shift may have primed the participants to respond to questions with the cued item as the salient object in their working memory, leading the children in Nappa et al.'s (2009) task and the adults in Staudte and Crocker's (2011) task to treat the gazed-at target as the object of the sentence.

Nappa and Arnold (2014) addressed this possibility with a series of experiments using videos showing the first author gazing or gazing and pointing at one of two characters (toys) or a neutral central object. In the first experiment, while viewing these videos participants heard short stories about the characters in which the final sentence was ambiguous, such as “Puppy is having some pizza with Panda Bear. He wants the pepperoni slice” (Nappa & Arnold, 2014, p. 64). Participants were asked questions about these stories after each trial. In experimental trials, the question was always about the identity of the actor (“He”) in the final sentence. When the gaze and pointing cues were neutral, participants showed a tendency to identify the actor as the first character mentioned in the previous sentence (N1). However, when gaze was allocated to the second character (N2), significantly more participants identified the actor as the second character and this effect increased when pointing cues were involved. These results are in line with those of Staudte & Crocker (2011), but still do not discount the possibility that low-level attentional processes are causing this effect. In their second experiment, Nappa and Arnold (2014) used the pointing and gaze cue condition from the first experiment and compared this to a new condition in which gaze and pointing cues were replaced with a flashing attention-capture stimulus (a solid black square) over one of the characters. When this stimulus was placed over N2, response time was significantly slower, suggesting that attention was captured by the stimulus. However, the location of the attention-capture stimulus was found to have no effect on the participant responses to the experimental question. This finding provides strong evidence that the preference for referring to a speaker’s previously gazed-at locations when a listener produces new sentences is not driven by low-level attentional processes, but instead by the listener’s understanding of the speaker’s communicative intentions.

While the above studies demonstrate the effect of gaze cues on language comprehension, Knoeferle and Kreysa (2012) demonstrated the effect of altering spoken language on the utilisation of gaze cues. Participants in this experiment heard spoken sentences while viewing a visual scene on a computer monitor. The visual scenes were videos featuring a computer monitor with three avatars from the virtual world “Second Life”. The sentences described the avatars on the screen and were either presented in the common German subject-verb-object (SVO) structure or an uncommon (yet still grammatically legal) object-verb-subject (OVS) structure. The experimenters manipulated whether or not the videos also contained a person watching the screen contained in the visual scene. The person in this video looked from the first to the second referent avatar at the onset of the verb in the sentence. The experiments found a facilitative effect of this gaze cue on sentence comprehension, but this facilitation was far more pronounced when the gaze cue accompanied the common SVO sentences compared to the less common OVS sentences. Knoeferle and Kreysa (2012) proposed that the additional processing difficulty of OVS sentences left fewer cognitive resources available to utilize the gaze cues, suggesting that the more difficult language is to process, the less listeners make use of the speaker’s gaze cues.

In Knoeferle and Kreysa’s (2012) task, all of the information required to understand the sentence was contained in the spoken language, so the gaze cues were purely supportive to language. The studies investigating natural communication and gaze cues that are outlined earlier in this section found that listener’s can identify a target using gaze cues before language unambiguously describes the target (Hanna & Brennan, 2007) and that speakers will sometimes change their utterances to exclude a verbal reference if the non-verbal cues of a listener show the listener to already be attending the referent item (Clark & Krych, 2004). Given this evidence that ambiguous language occurs naturally in conversations and that gaze cues are used to disambiguate these, Macdonald & Tatler (2013) investigated the interaction between language and gaze cues using unambiguous and ambiguous instructional sentences.

The paradigm used by Macdonald and Tatler (2013) considered gaze seeking and following behaviour in a real world communicative task. This involved a one-to-one interaction between an instructor (the experimenter) and a participant, in which the instructor directed the participant to build abstract structures out of building blocks. In the task, experimental manipulations were made during instructions that referred to block-selection. The instructor varied whether gaze cues were present or absent as well as the specificity of his block-selection instructions. Participants were found to only seek and follow gaze cues when the language was ambiguous (it did not specify which single block the participant was meant to pick up). Furthermore, when language was ambiguous and no gaze cues were present, participants showed very little gaze seeking behaviour, showing clear evidence for a flexible use of gaze cues, dependent on the relative informativeness of language. It was also noted that even when gaze cues were necessary for the task and readily available, participants did not seek and follow these all of the time. This could potentially be due to methodological issues; the participants could not look at the blocks and the instructor's face at the same time. However, it was speculated (Macdonald & Tatler, 2013) that the same social factors affecting the results of Laidlaw et al (2011) were at play here. More specifically, the social cost of looking at the instructor frequently in each trial may have deterred participants from seeking and following these gaze cues.

Gaze cues have been shown to support language during natural communication by disambiguating linguistic uncertainties (Hanna & Brennan, 2007) and allowing a listener to infer the intentions of a speaker (Staudte & Crocker, 2011; Nappa & Arnold, 2014). As gaze cues are often used alongside language, it is clear that understanding the relationship between gaze and language is important to understanding how gaze cues are used during natural collaboration. There is also evidence that social presence can influence the extent to which people seek out useful cues (Macdonald & Tatler, 2013). For this reason, it is important that any investigation of how we use gaze cues should begin with observations of gaze behaviour in a natural interaction, in which spoken language and social context can potentially exert influence.

The present thesis

The literature discussed in the present chapter shows that the social context in which an interaction occurs affects the extent to which interactors look at the eyes and face of their partner (Argyle & Dean, 1965; Laidlaw et al., 2011) and the extent to which they follow their partner's gaze cues (Milgram et al., 1969; Gallup, Hale et al., 2012). Changes to social context that have been shown to affect these factors include the proximity of the interactors to each other (Argyle & Dean, 1965), the perception of the relative social status of the other interactor (Efran, 1968), the behaviour of the other interactor (Modigliani, 1971) and whether or not the other interactor is perceived as being an in- or out-group member (Liuzza et al., 2011). For these reasons, my investigation of gaze cueing and spoken language in natural interactions begins with a real-world eye tracking study, to find how people naturally use language and gaze to complete a collaborative task (Study 1). The participants either worked together completely freely, using only a recipe as a guide or they were given specific roles to fulfil that affected their social perceptions of their partner. Gaze and language behaviour were investigated in this task, as well as the influence of social perceptions on these factors

Other research in this chapter shows how the language used alongside a gaze cue can effect whether that cue is used or not (Knoeferle & Kreysa, 2012; Macdonald & Tatler, 2013). Gaze cues have been shown to speed up communication (Clark & Krych, 2004; Hanna & Brennan, 2007) and can be used to resolve ambiguities in language (Staudte & Crocker, 2011) by the listener inferring the intention of the speaker (Nappa & Arnold, 2014). The highly flexible and top-down use of gaze cues described by the language literature is somewhat at odds with the lab-based findings that gaze cues automatically capture attention. In addition to the largely observational Study 1, the current thesis uses controlled real-world and lab-based paradigms that are designed to be sensitive to any automatic gaze utilisation effects that may be present in natural collaborations (Studies 2-5).

Beyond considering automatic gaze cueing effects, the experiments in this thesis investigate whether the form of language used in an interaction, rather than the specificity of language (Macdonald & Tatler, 2013), influences gaze utilisation (Study 2). Furthermore, given the previous findings that the extent to which we follow gaze is affected by the cue reliability, the effects of varying the reliabilities of both gaze cues and language cues are explored (Studies 3-5).

Combined, the five studies in this thesis will provide new insights into the interplay between language and gaze cues in natural communication. The extent to which we use gaze during collaboration will be explored, as well as the influence that language and social context have on gaze cue utilisation.

Chapter 2:

Dual eye-tracking in a natural collaboration – Study 1

Introduction

People have a strong tendency to follow gaze cues (Friesen & Kingstone, 1998; Riciardelli et al., 2002) and to orient towards the eyes of others in social scenes (Birmingham et al., 2009). However, there is also evidence that the language and social context that accompany gaze cues can affect how people orient to and follow these cues (Knoeferle & Kreysa, 2012; Liuzza et al., 2011). In the real world, gaze cues are rarely used without these accompanying factors, yet most previous research has not considered them together. The aim of the present study was to measure eye movements in a real-world setting in order to observe how gaze cues are used spontaneously during a natural collaboration, in which spoken language is present. In order to consider the influence of social context on gaze and language, the social roles of the two individuals in the collaborative task were manipulated.

It is important to begin the investigation of gaze cues and language in collaboration with a naturalistic task, as there is evidence that social presence influences our eye movement behaviour. Despite the evidence that we have a tendency to fixate the faces and eyes of others in static social scenes (Birmingham et al., 2009), when others are actually present, we look at faces and eye much less (Laidlaw et al., 2011). Furthermore, although we know people orient attention in the direction of a gaze cue (Ricciardelli et al., 2002), in the real world people have been found to follow the gaze cues of another less when that other can observe them (Gallup, Hale et al., 2012). Therefore to understand how language and gaze work together in a natural interaction it is important to observe language and gaze in a real world context. There is inevitably a lack of control in real world tasks relative to lab-based paradigms, however the present thesis will follow the theoretical framework of cognitive ethology (Kingstone, Smilek & Eastwood, 2008). This framework involves observing behaviour in as ecologically valid an environment as possible and then using those findings to devise simpler experiments in less ecologically valid environments with fewer variables. Findings from the real world interactions in the present experiment will be used to inform the more controlled gaze cueing experiments in subsequent chapters.

In the present real world paradigm the onset of language and gaze cues could not be controlled or directed. Instead, natural behaviour was observed and coarse measures were adopted to indicate gaze seeking and following behaviour. In line with a method used by Macdonald and Tatler (2013), looks to the other person were used as an indicator of gaze seeking behaviour: the more one looked at their partner the more likely they were seeking information from the gaze of their partner. Additionally, this indicator was measured during instructional phrases, in order to investigate gaze-seeking behaviour when gaze cues were likely to be most helpful in the task. To investigate gaze following, we used shared gaze, which was defined as any instance in which participants looked at the same object. The initiator of each instance of shared gaze was also recorded as an indicator of who initiated each shared fixation. Gaze alignment has previously been shown to occur when participants work together on a task (Richardson et al., 2007) and has been shown to increase performance in visual search tasks (Brennan, Zelinsky, Hanna, & Savietta, 2012). If this alignment of gaze can help participants in computer-based tasks, then it is interesting to see if this alignment spontaneously occurs in real-world collaborations and, if so, for how long it occurs. These measures are rough indicators of the phenomena that are under investigation in this thesis, but the findings inform later and more controlled experiments.

Rather than being entirely observational, this dual eye-tracking study did involve one manipulation: participant pairs were either given social roles (a leader and a follower) or they were not. The manipulation of the perception of another has been used before in this research area. There is some evidence from early social interactions studies (outlined in detail in the previous chapter) that social perceptions can influence the extent to which we look at another, such as work by Efran (1968) that showed college Freshmen looked more at Seniors than fellow Freshman. More recently, participants carrying out a Posner (1980) task in Italy were shown distracter gaze cueing stimuli made from the faces of Italian political figures, including Silvio Berlusconi (Liuzza et al, 2011). The gaze of Berlusconi was found to cause significantly more interference in the task for in-group than out-group participants. Crosby et al. (2008) showed that participants were more likely to look at an individual on a monitor if they thought the individual could hear comments that were potentially offensive to that individual. These results show that social factors such as beliefs about another individual can affect how others look at them as well as how others look at external objects whilst communicating with them. Although these results show effects of prior beliefs about others on gaze behaviour, it is still unclear how beliefs about the role or knowledge of another affect the use of gaze cues in natural collaboration.

The present study aims to observe collaborations between two people in a natural environment. Participants, in pairs, were given a recipe to follow in order to make the batter for a cake. During this collaboration their eye movements were recorded using portable eye-trackers. This paradigm allowed us to observe natural eye movement behaviour during a natural collaboration featuring spoken language and social influences. When coding the data, the time participants spent looking at each other (partner gaze) or at the same object simultaneously (shared gaze) were of particular interest as well as the time taken for participants to look at an item after the onset of the verbal reference to that item. Although the collaboration was devised to be as natural as possible, social context was manipulated in one way: half of the pairs were given roles (chef or gatherer) to fulfil and the other half were not. By manipulating this, interactions between language, gaze and social context could be investigated. The hypothesis was that, in line with Efram's (1968) findings, the gatherers would look at their partner longer than other participants and would be less likely to initiate shared gaze, indicating a higher level of gaze seeking and following behaviour.

Method

Participants

Twenty-four undergraduate and post-graduate students from the University of Dundee participated in this study (Female = 16). They were split into twelve pairs to carry out the task. Six pairs were allocated to the roles condition and six were allocated to the no roles condition (see design). In the roles condition, three of the pairs were both female, one pair both male. Of the two mixed pairs, one had a female chef and the other a female gatherer. In the roles condition there were two all female pairs and four mixed pairs. Each individual was recruited separately, using the School of Psychology's participant recruitment system, and all participants were asked if they knew their partner. All but one pair (a mixed pair in the no roles condition) had never met their partner before. Undergraduate students received course credits for participating.

Materials

The experiment took place in a kitchen area on the University of Dundee campus. The kitchen was fully equipped with standard kitchen appliances, but only the oven and microwave were used. All items and foodstuffs that could be removed were removed before testing and the experimental materials were arranged carefully around the kitchen. This included the items and foodstuffs that were to be used for the procedure (a knife, a spoon, a set of kitchen scales, a large bowl, a small bowl, butter, sugar, whisk, eggs, sieve, flour, milk, and a square tin) as well as a selection of distractor items (a jar of instant coffee, a bottle of table sauce, a jar of cooking sauce, a bottle of olive oil, a bottle of wine, a salt shaker, a box of tea bags, an insulated coffee cup, a box of muesli and an empty re-usable shopping bag). All of these items were placed in the same location for each pair of participants. A Recipe Procedure sheet was provided for each pair (*Appendix 3*). This sheet explained, step-by-step, how to make the batter for a Victoria Sponge. There was also a Chef Guidelines sheet (*Appendix 4*) and a Gatherer Guidelines sheet (*Appendix 5*) for those in the roles condition. These sheets explained the responsibilities and duties for participants in the chef and gatherer roles. Eight toy building blocks (Megabloks) were required to calibrate the eye-trackers.

Design

This experiment had a between subjects design. The two independent variables for the analysis of shared gaze and the time participants spent looking at each other (partner gaze) were the use of roles (roles or no roles) and the allocation of roles within the roles condition (chef or gatherer). For the analysis of the references used in the instructions, the independent variables were the use of roles (roles or no roles) and identity of participant (speaker or listener).

Procedure

This experiment required two participants to work together. The experimenter began by fitting a portable eye tracker to the first participant outside the kitchen. At this point in the roles condition the first participant was given the Chef Guidelines and the second participant was given the Gatherer Guidelines. They were both instructed to read over their sheet and make sure they understood their roles. The Chef Guidelines informed the chef that they were in charge of preparing the recipe and that the gatherer was there to assist them. The sheet explained that the chef was expected to mix and prepare ingredients, following a recipe that they could not show to the gatherer. The chef would not be expected to collect any items or foodstuffs, but to delegate those duties to the gatherer. The chef would also be able to ask the gatherer to assist them with any aspect of the preparation they wished. The Gatherer Guidelines explained that the gatherer would not be expected to make any decisions concerning the preparation, but should instead do as instructed by the chef. Once the participants declared they understood their roles the gatherer was asked to remain outside whilst the experimenter and the chef entered the kitchen. The experimenter then gave the chef the Recipe Procedure sheet and told the chef where all of the necessary items and foodstuffs were located. The chef was then told they would have approximately three minutes to familiarise themselves with the kitchen and the locations of the items. During these three minutes the experimenter fitted another portable eye-tracker to the gatherer. In the no roles condition the second eye-tracker was fitted straight after the first. At this point, in both conditions, the participants were brought into the kitchen and the eye-trackers were switched on.

The participants were instructed to stand facing each other with their eyes closed. A flash of light was then emitted from a camera flash in between the two participants. This flash was picked up by all four cameras across the two eye-trackers and was later used as a reference point to synchronise all of the videos. The first stage of calibration then took place for each participant in turn using the Megabloks in a circular array on a kitchen counter. This was followed by a second stage of calibration on a vertical plane using reference points on the experimenter's body at a distance of about three metres (see *Eye movements and sound recording* section for more details). Once this stage of calibration was complete those in the no roles condition were directed to the Recipe Procedure sheet and informed that all of the items they would require were located around the kitchen. All participants were informed that the experimenter would be standing outside the kitchen, out of sight and that the participants must make no attempt to interact with him. The experimenter then told the participants that they should begin as soon as he was out of the room. The experimenter left and the procedure began. Once the participants had put the batter mixture in the oven (the final step) the experimenter entered and the earlier procedure with the camera flash was repeated. The eye-trackers were then switched off and removed and the eye-movement and audio data were later analysed.

Eye movement and sound recording

Participants' eye movements were tracked using Positive Science LLC mobile eye trackers, which allowed free head movement. Each eye tracker has two cameras mounted on the frame of a pair of spectacles: one records the scene from the participant's point of view and the other records the right eye. Data from these cameras were captured on digital camcorders (Figure 2.1). For one of the eye-trackers these camcorders were stored, alongside a power supply for the eye-tracker, in a lumbar pack worn by the participant. The camcorders connected to the second tracker were again stored alongside a power supply, but were stored in a light backpack worn by the participant. This eye tracker also has a small microphone attached to the frame. This microphone recorded sound throughout the experiment and was able to pick-up the voices of both participants. Gaze direction was estimated off-line using Yarbus software provided by Positive Science, LLC, which tracks the pupil and corneal reflection. Calibration was carried out in two stages, one looking down at a counter and the other looking across the room. These two stages were used because by tracking one eye, the vergence of the eyes that occurs as participants focus on objects at different distances cannot be directly measured. Instead the model was fit to fixations on both proximal and distal points. The first stage involved the participant fixating on Megabloks in a circular-array as the experimenter pointed to them. The second stage involved the participants, at a distance of approximately three metres, fixating points on the experimenter's body as he called them out. These included the tip of the experimenter's left index finger as it pointed up and to the left and the opposite for his right index finger, the experimenter's nose, the tips of the experimenter's left and right shoes and finally the top of the experimenter's thumb whilst it was positioned centrally over his abdomen. If the tracker estimates in the scene video fell on the correct blocks and correct body positions the calibration was deemed adequate. Eye movement data were recorded at 30Hz with a spatial accuracy of about 1 degree. Once videos for both participants were rendered with the eye movement information, Quicktime Pro was used to synchronise both videos in to one movie file, ready for analysis.

Figure 2.1. Two still images from the rendered eye tracker movies. The top still shows an example of *partner gaze*, with person B (left) looking at person A. The bottom still is from the same pair 19 frames (640ms) later and shows an example of *shared gaze*, with both participants looking at a plastic spoon.

Video coding and dependent variables

Eye tracking data were coded manually offline using Quicktime Media Player and audio information was extracted using Audacity sound editing software. The first three dependent variables investigated when participants looked at the same objects (shared gaze). These variables were (1) the proportion of task time in which shared gaze occurred, (2) the proportion of shared gaze events initiated by person A and (3) the time difference between the start of a shared gaze event and time the initiator first fixated the object. For these analyses, in each pair, one participant was labelled person A and the other was labelled person B. In the roles condition person A was the chef and B was the gatherer. Since there were not any defined roles in the no roles condition, participants in this condition were arbitrarily allocated as person A or B. The coding of these data was split between the lead experimenter and three undergraduate volunteers from the School of Psychology. The lead experimenter provided thorough written guidelines for coding for the volunteers and provided them with the opportunity to ask questions at any time. To begin, all four coders coded the same movie file and these were all compared by the lead experimenter to ensure a consistent and high quality of coding. All coders went through the videos frame-by-frame and marked down frame references for the beginning and end of each shared gaze event. Additionally the initiator of each shared gaze event was noted, along with the frame references for the time the initiator began fixating on the shared attention location.

The video coders also noted the times the participants looked at each other (partner gaze) throughout the task. These data were used for the first two partner gaze measures: (4) the proportion of task time spent on partner gaze and (5) the mean length of partner gaze events. The remaining two partner gaze measures involved additional coding. Individual instructions were coded and analysed. The lead experimenter alone coded these instructions, using Audacity sound editing software and the Quicktime movie files. For each pair, each instruction statement was numbered and transcribed, noting the speaker. The start time and end time (to the nearest frame reference) of each instruction was coded, along with the time that the speaker first looked (if at all) at the listener, the time the listener first looked at the speaker and whether or not the speaker used spatial language or pointed during their instruction. In the roles condition, the speaker was always the chef and the listener always the gatherer. In the no roles condition the participant who gave the instruction was considered to be the speaker. Therefore the identity of the speaker and listener would switch throughout each movie in the no roles condition. From coding these data, two dependent variables for the analysis of partner gaze were used: (6) the proportion of instructions in which one or more partner gaze event occurred and when these events occurred (7) the proportion of partner gaze events that involved the listener looking at the speaker.

During instruction coding, individual references to objects were also coded. Each instruction had between one and four references to relevant objects. Like the instructions, each reference was numbered and the start and end times were recorded. The first time the speaker and listener fixated on each referent object was coded as well as whether the reference was made using the name of the object, a pronoun or if it wasn't directly mentioned verbally. This coding allowed for the investigation of the effect that a number of different factors had on (8) the time difference between the fixation on a referent object and the onset of the verbal reference relating to that object. These factors included person (speaker or listener) and the presence of roles, but also type of language used. Previous research of natural dialogue has shown that speakers will align speech with their partners (Pickering & Garrod, 2004). This alignment can involve ambiguous utterances (such as those using pronouns rather than proper nouns) being interpreted unambiguously, due to the context provided by previous utterances or the constraints of a shared visual environment (Brown-Schmidt & Tanenhaus, 2008). The time between the onset of a reference phrase and a listener's fixation to the referent object can provide an indicator of alignment between interactors. The present study will investigate any differences in this measure found between different types of verbal reference (proper noun or pronoun and spatial or non-spatial).

Statistical Analysis

For some of the time measures, the distribution of data was skewed to lower values. In these cases the data were $\log_{10}$-transformed. When time difference measures involved negative values, a constant was added to these values before $\log_{10}$-transformation to make all values positive. An independent-samples t-test was used to compare the proportion of task time in which shared gaze occurred between the roles and no roles condition and a 2 (roles condition) × 2 (participant) ANOVA was carried out on the proportion of task time spent engaged in partner gaze. For all other analyses the *lme4* package (Bates, Maechler, Bolker & Walker, 2014) in the R statistical programming environment was used to run linear mixed effects models (LMMs). Many of the measures used in this study have uneven distributions of data, making traditional ANOVA models inappropriate for these measures. The details of the approach to LMMs used in this study are outlined in this section.

Random Factors. Typically, LMMs use subject and item as random factors, providing a benefit over AVOVAs in which subject and item analysis must be carried out separately. In the present study, there was no item equivalent and the task was carried out in pairs, so subject pair was the sole random factor. Following the guidance of Barr, Levy, Scheepers and Tily (2013) random slope models were used, allowing the models to consider the different effects that the fixed factors have on different subject pairs. Models that are maximal were used where possible.

Simplifying models. The maximal (most complex) models can often fail to converge if there are not enough observations in the data. In these cases, simpler models were attempted. Models were simplified, when necessary, by first removing correlations between random-slopes and intercepts. After this, the slope of any interaction from the random factor was removed. If the model still failed to converge, the intercept was removed, followed by a fixed factor slope. The most complex model that successfully converged was always used (See *Appendix 1* for more details). Throughout this thesis LMMs are numbered and the full models can be seen in *Appendix 2*.

Sum coding. All of the fixed factors in these LMMs were categorical variables, therefore they required coding. A form of effects coding known as sum coding was carried out. This method involved initially dummy coding a factor with two levels, "0" for one level and "1" for the other. After dummy coding the mean value for this variable was subtracted from each case. After this process the resulting variable would always have a mean of "0", even if the factors were unbalanced.

Calculating *p*-values. To calculate the p-value for an effect of any given fixed factor, a second model was produced, which was identical to the original LMM except for the omission of the fixed factor under investigation. This model was then compared to the original model using the *anova()* function in the *car* library (Fox & Weisberg, 2011). The *p* value from these model comparisons was used as the *p*-value for the effect of the fixed factor.

Pairwise comparisons. When significant or approaching significant interactions were identified, post-hoc comparisons were carried out with Tukey tests using the *glht()* function in the *multcomp* library (Hothorn, Bretz & Westfall, 2008).

Results

Shared gaze analysis

Shared gaze was defined in this experiment as any instance in which both participants were fixating the same object. Shared gaze was regarded as an indicator of shared attention in the task. Figure 2.2a) shows the mean proportion of task time in which shared gaze occurred for subject pairs in the roles and no roles conditions. Subject pairs in the roles condition spent a higher mean proportion of time on shared attention (.26) than pairs in the no roles condition (.21), but this difference was not found to be significant, $t(10) = 1.286$, $p = .228$.

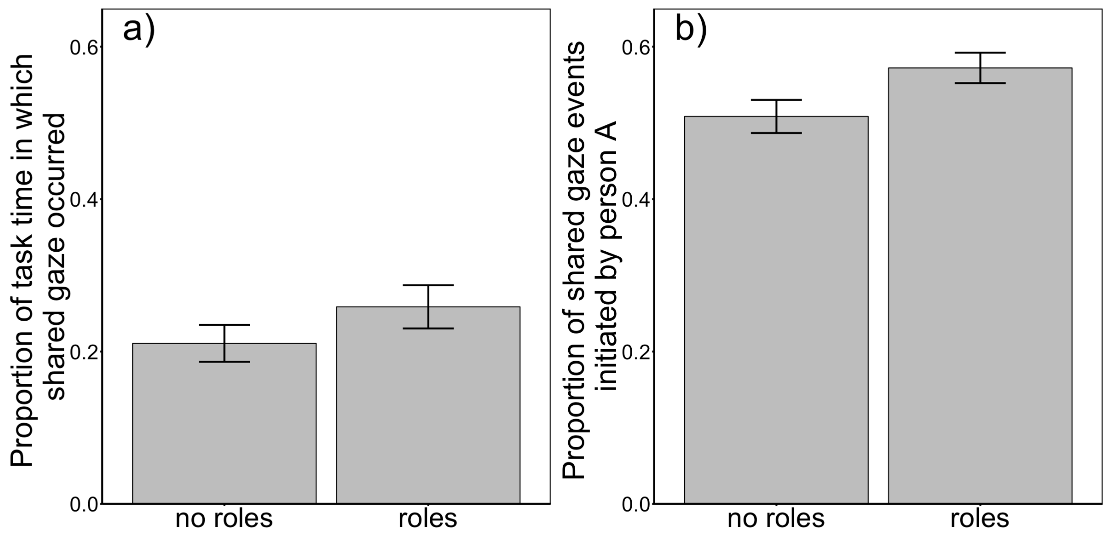

Figure 2.2. The (a) Proportion of task time in which shared gaze occurred and (b) the proportion of shared gaze events initiated by person A in the no roles and roles conditions. Standard error is shown in both graphs. Mean task time was 9 minutes 19 seconds.

Figure 2.2b) shows the proportion of shared gaze events that were initiated by person A in each of the roles conditions. In the no roles condition, where letter assignment was arbitrary, person A initiated about half of the shared gaze events (.51). In the roles condition, where person A was always the chef, person A initiated more of the shared gaze events (.57). However, a random-slope LMM (Model 2.1) of shared gaze event initiator, using roles condition as a fixed factor and subject pair as a random factor, found that roles condition had no significant effect on the proportion of shared gaze events initiated by person A, $\beta = .211$, $SE = .39$, $z = .557$, $p = .591$.

To measure the mean time taken for an instance of shared attention to occur, the difference between the time the initiator looked at any object and the time that shared gaze on that object began was calculated. Figure 2.3 shows the mean values of this time difference for both person A and person B in the roles and no roles condition.

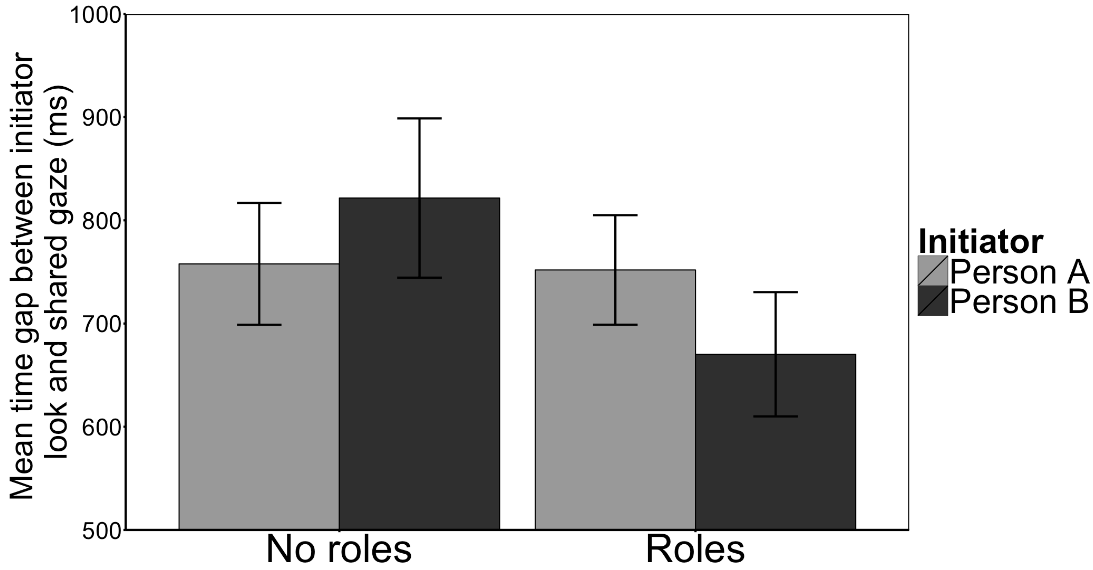

Figure 2.3. The mean time gap between the beginning of the initiator's fixation to shared gaze location and the initiation of shared gaze. Results are shown (with standard error) for shared gaze events initiated by person A and person B in the no roles and roles conditions.

Figure 2.3 shows that the mean time gap is smaller in the roles condition for both person A (25.0 frames) and person B (21.4 frames) than in the no roles condition (person A = 31.0 frames, person B = 36.8 frames). A random slope LMM (Model 2.2) of the $\log_{10}$ time difference was carried out, using presence of roles, initiator and their interaction as fixed factors and subject pair as a random factor. The model showed that the time differences were significantly smaller in the roles condition compared to the no roles condition, $\beta = -.097$, $SE = .049$, $t = -1.980$, $p = .048$. Figure 2.3 also shows there was a difference in mean time difference between person A and person B in the roles but not the no roles condition. However, the LMM showed that there was no significant effect of initiator, β = .036, *SE* = .045, t = .800, p = .365, nor any interaction, β = .083, *SE* = .091, t = .920, p = .328.

Partner gaze analysis

Any instance in which one participant looked at the other in the task is referred to as partner gaze. The proportion of task time in which partner gaze occurred is shown in Figure 2.4.

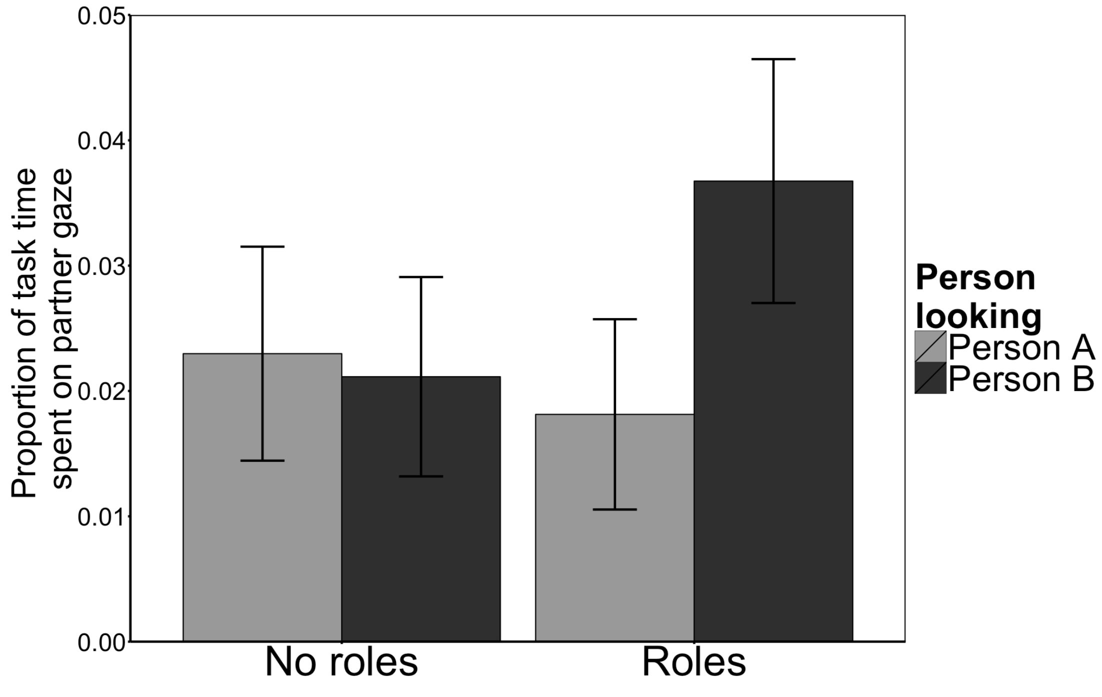

Figure 2.4. The proportion of task time in which partner gaze occurred for person A and person B in the no roles and roles conditions. Error bars show standard error. Mean task time was 9 minutes 19 seconds.

The proportion of task time spent on partner gaze in the roles condition was higher for person B (.037) than person A (.018), but in the no roles condition the proportion for person A (.023) and person B (.021) were similar. A 2×2 ANOVA found no significant main effect of roles condition, $F(1,20) = .402$, $p = .533$, or person looking, $F(1,20) = .976$, $p = .335$, nor any significant interaction, $F(1,20) = 1.451$, $p = .242$.

The next partner gaze measure was the mean time taken for instances of partner gaze for person A and person B in the no roles and roles condition. The mean times of partner gaze events are shown in Figure 2.5.

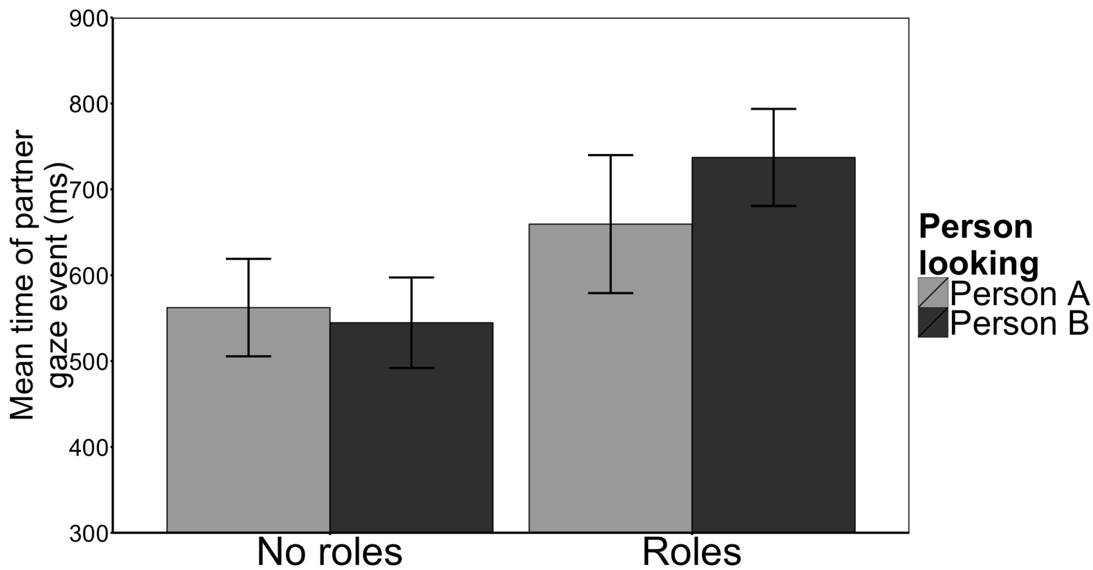

Figure 2.5. The mean duration of partner gaze events. Results are shown (with standard error) for partner gaze events in which person A looked at person B and person B looked at person A in the no roles and roles conditions.

The mean duration of partner gaze was found to be higher in the roles condition for both person A (19.8 frames) and person B (22.1 frames) than in the no roles condition (person A = 16.9 frames, person B = 16.3 frames). A random slope LMM (Model 2.3) of the $\log_{10}$ partner gaze time was carried out, using presence of roles, person and their interaction as fixed factors and subject pair as a random factor. A significant effect of roles was found, $\beta = .119$, $SE = .053$, $t = -2.240$, $p = .036$. There was no significant effects of person, $\beta = -.026$, $SE = .075$, $t = -.350$, $p = .696$, nor was there a significant interaction, $\beta = -.032$, $SE = .150$, $t = -.210$, $p = .863$.

Partner gaze was also analysed specifically for times in which one participant was giving the other participant an instruction. Because these instances of partner gaze are coupled with verbal instructions, these instances are likely to be related to communication. Figure 2.6a) shows the proportion of instructions in which at least one instance of partner gaze occurred for the roles and no roles condition. A random-slope LMM (Model 2.4), using roles as a fixed factor and subject pair as a random factor, showed that the proportion of instructions with partner gaze was not significantly different between the no roles (.32) and roles (.51) conditions, $\beta = -.808$, $SE = .568$, $z = 1.423$, $p = .155$.

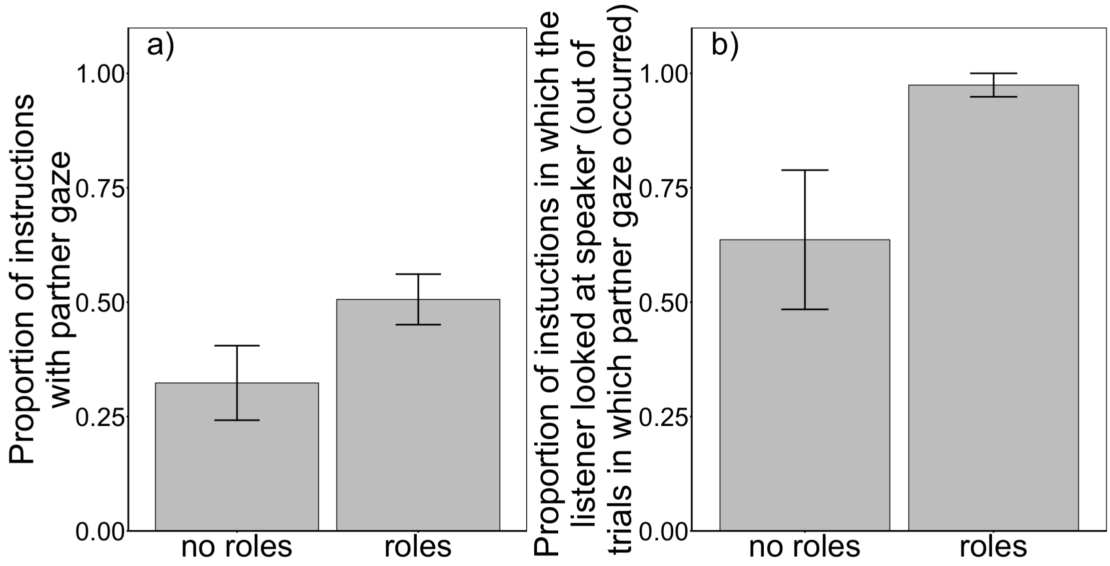

Figure 2.6. The a) Proportion of instructions in which partner gaze occurred and b) the proportion of instructions with partner gaze in which the listener looked at speaker in the no roles and roles conditions. Standard error is shown in both graphs.

Figure 2.6b) shows that the proportion of instructions with partner gaze in which the listener looked at the speaker was higher in the roles condition (.97) than the no roles condition (.64). A random-slope LMM (Model 2.5), using presence of roles as a fixed factor and subject pair as a random factor, showed the difference between roles to be significant, $\beta = 2.983$, $SE = 1.348$, $z = 2.214$, $p = .027$.

Reference analysis

The time difference between the onset of a verbal reference and time the participants fixated the referent item provides insight into how the participants engaged with the instructions. The first reference measure compared this time difference across participant and roles conditions. Rather than assigning participants as person A or person B, for each reference in these analyses one participant was labelled as the speaker of the reference and the other was labelled as the listener of the reference. In the roles condition, the chef was always the speaker, however in the no roles condition the identity of the speaker varied between references. Data were not normally distributed and contained negative values. Means are shown in Figure 2.7, with positive values indicating the verbal references came before the fixations on the referent objects. For the purposes of the LMM analysis, a constant of 106 frames was added to the data, which was then $\log_{10}$ transformed.

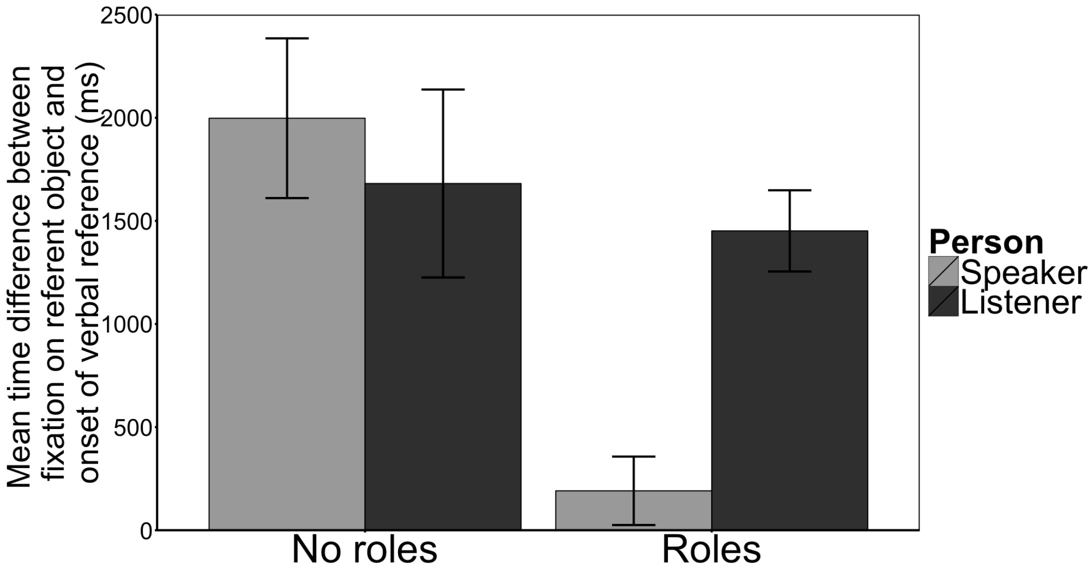

Figure 2.7. The mean time difference between referent object fixation and the onset of the verbal reference. Results are shown (with standard error) for the speaker and listener of the reference in the no roles and roles conditions.

The mean time difference between referent object fixation and the onset of the verbal reference was similar for the both the speaker (59.9 frames) and listener (50.4 frames) in the no roles condition. However, in the roles condition the $\log_{10}$ time difference was smaller for the speaker (5.7 frames) than the listener (43.5 frames). A random-slope LMM (Model 2.6) of the $\log_{10}$ time difference was carried out, using person, role condition and their interaction as fixed factors and subject pair as a random factor. There was a significant effect of both participant, β = -0.084, *SE* =.022, *t* = -3.720, *p* < .001, and role condition, β = -.095, *SE* =.042, *t* = -2.250, *p* = .030, as well as a significant interaction, β =-.121, *SE* =.049, *t* = -2.510, *p* = .012. A post-hoc Tukey test showed a significant difference between speaker and listener in the roles condition (p<.001), but not the no roles condition (p >.999).

Whether the use of the object name as a reference rather than a pronoun had any effect on the time difference between fixation on the referent object and onset of the verbal reference was investigated. Because data were not normally distributed and contained negative values, a constant of 106 (for speakers) and 87 (for listeners) was added to the data, which were then $\log_{10}$ transformed. Figure 2.8 shows the (untransformed) mean time difference for references using object names and pronouns in the roles and no roles conditions for a) speakers and b) listeners. The mean time difference for speakers was larger for object name references in the no roles condition (63.6 frames) compared to the roles condition (1.5 frames). For the pronoun references there was a much smaller difference between the no roles (29.5 frames) and roles conditions (18.3 frames). A random-slope LMM (Model 2.7) of the $\log_{10}$ transformed data, using roles, reference type and their interaction as fixed factors and subject pair as a random factor, confirmed a significant interaction, β =-.197, *SE* =.091, *t* = -2.150, *p* = .042, as well as a significant effect of roles condition, β =-.163, *SE* =.048, *t* = -3.430, *p* = .003. There was no effect of reference type, β =-.005, *SE* =.042, *t* = -.120, *p* = .963. A post-hoc Tukey test showed that the difference between results in the no roles and roles condition was significant for object name references (p <.001) but not pronoun references (p >.999).

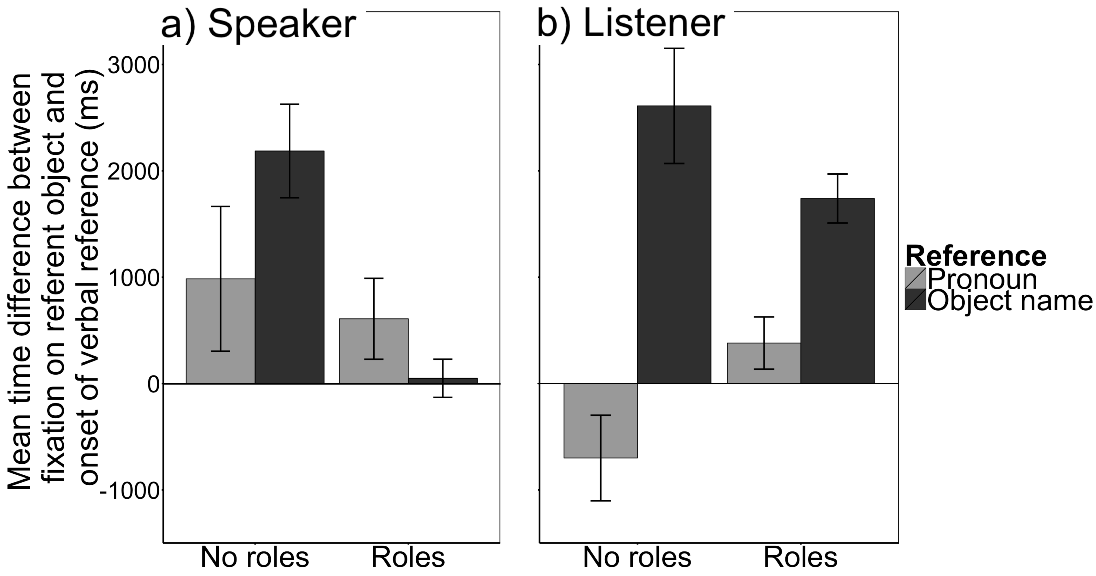

Figure 2.8. The mean time difference between referent object fixation and the onset of the verbal reference for a) speakers and b) listeners. Results are shown (with standard error) for the pronoun and object name references in the no roles and roles conditions.

Figure 2.8b) shows that time difference between referent object fixation and the onset of the verbal reference for listeners was higher for object name references in the no roles (78.3 frames) and roles (52.2 frames) conditions than pronoun references (no roles = -21.0 frames, roles= 11.4 frames). A random-slope LMM (Model 2.8), using roles and reference type and their interaction as fixed factors and subject pair as a random factor, showed that reference type had a significant effect on the time difference, β =.206, *SE* =.047, t = 4.410, p < .001. There was no effect of roles condition, β =-.044, *SE* =.048, t = -.920, p = .313, but there was a significant interaction, β =-.212, *SE* =.105, t = -2.020, p = .040. A post-hoc Tukey test showed the difference between object name and pronoun references to be significant both in the no roles ($p<.001$) and roles conditions ($p = .038$).

The mean time difference was also compared between the referent object fixation and the onset of the verbal reference between references provided with and without spatial information. Again, data were not normally distributed and contained negative values, therefore a constant of 73 (for speakers) and 67 (for listeners) was added to the data, which was then $\log_{10}$ transformed. Figure 2.9 shows the (untransformed) mean time difference for these references in the roles and no roles conditions for a) speakers and b) listeners, with positive values indicating that the onset of the verbal reference preceded the fixation to that reference. The mean time difference was higher for the speaker in the no roles condition for both non-spatial (71.4 frames) and spatial (57.3 frames) references than the roles condition (non-spatial = 9.7 frames, spatial = 10.3 frames). A random-slope LMM (Model 2.9) of the $\log_{10}$ transformed time difference, using roles, reference type and their interaction as fixed factors and subject pair as a random factor, showed a significant effect of roles condition, β =-.249, *SE* =.073, t = -3.420, p = .002. There was no effect of reference type, β =-.038, *SE* =.049, t = .770, p = .468, nor was there a significant interaction, β =-.154, *SE* =.107, t = -1. 440, p = .210.

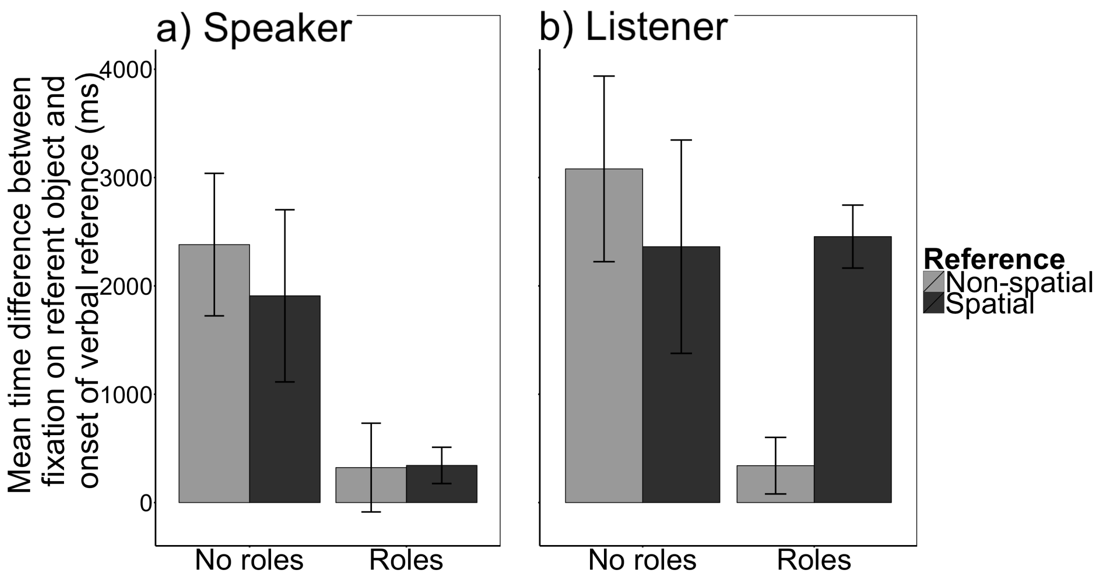

Figure 2.9. The mean time difference between referent object fixation and the onset of the verbal reference for a) speakers and b) listeners. Results are shown (with error bars) for references provided alongside non-spatial and spatial information in the no roles and roles conditions.

Figure 2.9b) shows that mean time difference between referent object fixation and the onset of the verbal reference for listeners varies for non-spatial references between the no roles (92.4 frames) and roles (10.2 frames) conditions, but not for spatial references (no roles = 70.9 frames, roles= 73.7 frames). A random-slope LMM (model 2.10) of the $\log_{10}$ transformed time difference, using roles, reference type and their interaction as fixed factors and subject pair as a random factor, showed an interaction to be approaching significance, β =.225, *SE* =.119, t = 1.900, p = .052. The difference across role conditions did reach significance, β =-.131, *SE* =.068, t = -1.920, p = .049, and the effect of reference type was clearly significant, β =.186, *SE* =.050, t = 3.750, p = .002. A post-hoc Tukey test showed the difference between the non-spatial and spatial reference conditions to be significant in the roles condition ($p<.001$) but not the no roles condition ($p = .999$).

Given the similarity in the effects of spatial language and object names on the time difference between referent object fixation and the onset of the verbal reference, the relationship between spatial language and the use of object name references was investigated. Figure 2.10 shows the proportion of references using spatial language for pronoun references and object name references.

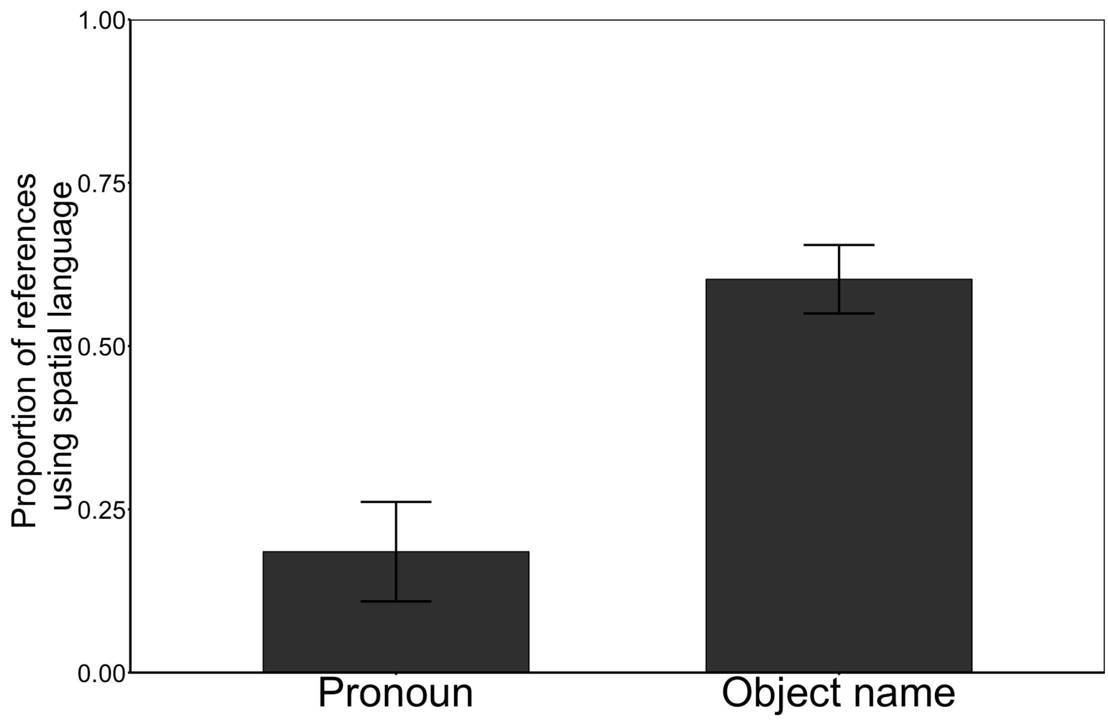

Figure 2.10. The proportion of references provided alongside spatial information (with standard error shown) for references using a pronoun or the name of the referent object.

A higher proportion of object name references was supplied with spatial references (.60) compared to pronoun references (.19). A random-slope LMM (Model 2.11), using reference type as a fixed factor and subject pair as a random factor, showed that this difference was approaching significance, β =2.581, *SE* =1.513, z = 1.706, p = .088.

Discussion

The present study explored the effects of manipulating the presence of social roles on gaze and language in a real world collaborative task. The presence of roles in collaboration influenced how participants engaged in shared gaze as well as partner gaze. The latter was also affected by the particular role of the individual within each pair. The use of proper nouns and spatial language influenced the time taken for participants to gaze at referent locations and these influences were mediated by the presence of roles and the identity of the participant. The results of this largely exploratory study provide evidence that the presence of social roles affects the extent to which participants visually engage with a collaborator and their surroundings. Furthermore, the presence of roles affected the relationship between language-type and gaze allocation. These findings provide the foundation for later, more controlled experiments on the interaction between gaze cueing and language.

There was no significant difference between the proportion of time that partner gaze occurred across roles conditions. Participants spent far less time (between 2-4%) looking at each other than they spent on shared gaze (between 20-25%). These results appear to be at odds with the results of some previous lab-based studies. People have been shown to have a preference for looking at eyes when viewing pictures of people

(Yarbus, 1967) or social scenes (Birmingham et al, 2009; Zwickel & Võ, 2010), however in this task participants spent very little time looking at their partners. Given the potential informativeness of the eyes (Tomasello et al, 2007) and the ease with which people can interpret gaze direction (Anderson et al. 2011) this finding may seem surprising. However, studies using real people as stimuli may offer an explanation. Laidlaw et al. (2011) showed that people were less likely to look at a present confederate than the same confederate on a video monitor and Gallup, Hale et al (2012) found that people were less likely to follow the gaze of strangers who could see them than strangers who could not. The authors of these studies concluded that this was due to there being potential consequences (social interaction) to looking at the present confederate or the on-coming stranger. The present study differed from these other real world studies in one important aspect: the participants were already involved in an interaction. The low proportion of time spent on partner gaze suggests that the social influences that dissuade people from looking at strangers also have an effect during an interaction. This tendency to avoid long looks to a collaborator can therefore not simply be explained by the wish to avoid a social interaction, as the interaction was already initiated. To speculate, it may be that participants avoided spending long looking at each other in order to follow social norms or it may simply be that in this task, there were more useful places to look than the collaborator.

These results present an obvious question; if people rarely look at each other in an interaction, can they still utilise gaze cues? Although the results cannot provide a definitive answer, there are three main arguments for the ability to utilise gaze cues in these circumstances. Firstly, it has been shown that gaze cues can be followed and affect language comprehension, even when they are not directly fixated (Knoeferle & Kreysa, 2012). Secondly, successfully identifying a gaze cue may not need a large overall proportion of time spent on partner gaze, but may require effective individual partner gaze events. Therefore instead of spending a long time looking at the eyes of another, an effective way to follow gaze cues may be to look to the eyes briefly and infrequently at appropriate times. Thirdly, it may be the case that eyes are generally not sought out during a task, but are used when required, for example, during instructions.

One way to investigate the second possibility is to consider the features of a partner gaze event that may make the event more effective at perceiving nonverbal signals, such as gaze cues. One feature that could have reasonably influenced effectiveness in this way is the length of individual partner gaze events. In this study, the mean length of partner gaze events was found to be significantly longer in the roles condition than the no roles condition. This may indicate that participants in the roles condition were more effectively attending to the non-verbal cues of their partners, resulting in the shorter gaps found between initiator fixation time and onset of shared attention in the roles condition.

To consider the third possibility, partner gaze was measured when the gaze allocation of a partner may have been particularly informative: during spoken instructions. Participants in the roles condition were no more likely to look at their partner than those in the no roles condition, but when there were instances of partner gaze during an instruction, listeners looked at the speaker significantly more in the roles condition compared to the no roles condition. This finding shows that our preference for looking at others can be affected by social context. In the roles condition the listener was always the gatherer, following instructions given by an informed chef, who was in charge. In the no roles condition the identity of the listener would switch between the two equal partners, depending on who was giving the instruction. Macdonald and

Tatler (2013) found that the degree of informativeness of gaze cues affected the extent to which the cues were sought out, with highly informative cues being sought most often. One possible interpretation of the present findings could be that the manipulation of the roles of the participants effectively manipulated the perceived informativeness of the cues provided by the chef: listeners in the roles condition may have considered the gaze cues of the chef to be highly informative, whereas the gaze cues of the speaker in the no roles condition may have been considered less informative.

The observed pattern of results may not have arisen from an effect of perceived informativeness, but rather from a social effect of authority. Liuzza et al. (2011) found that right-wing voters were more heavily influenced by the gaze cues of their political leader than the gaze cues of the opposition leader. In the roles condition, the chef was in charge of the procedure and was therefore the leader of the gatherer. It is possible that, as well as being more inclined to follow the gaze cues of a leader, people are also more inclined to orient to the leader's gaze cues. Previous research has shown a fixation-bias for higher social-status individuals when viewing videos (Foulsham, Cheng, Tracy, Henrich, & Kingstone, 2010) and the present results extend these findings to a real-world interaction. Although the results do not allow us to favour one explanation over the other, these findings provide good evidence that the social context of collaboration can affect the extent to which collaborators look at each other during communication. A more controlled future experiment may be able to distinguish between the effects of the perceived reliability of a person and the perceived social role of a person.

As with partner gaze, the overall time participant pairs fixated the same object (shared gaze) was not found to vary significantly across roles condition. It was expected that more shared gaze events would be initiated by person A in the roles condition because person A was the leader (chef) of the pair and initiating a shared gaze event was an indicator of providing a gaze cue that was followed. However this difference was not found to be significant. The mean time difference between the initiator of a shared gaze event fixating the gazed-at item and the initiation of the shared gaze event was significantly shorter for pairs in the roles condition compared to those in the no roles condition. The cause of this difference is unclear, but to speculate, one explanation is that the presence of distinct social roles increased the extent to which each participant attended to the other. Early research on non-verbal communication in real social interactions found that dominant interactors looked more at partners when talking and subordinate interactors looked more at their partners when listening, perhaps because the dominant interactor wanted to ensure they were being listened to and the subordinate wanted to signal that they were listening (Ellyson et al., 1981). It may be the case that those in the roles condition attended to their partners more and this led to them responding more quickly to the gaze location of their partners.

The relationships between the language used in verbal instructions, gaze allocation and social roles were also investigated. The time difference between the onset of a verbal reference and time the participants fixated the referent item was the measure for this part of the analysis. Overall, speakers in the roles condition were the quickest to fixate the referent item after the onset of the verbal reference, likely due to the chefs knowing where the objects were located before the procedure began. Participants either used the proper name or a pronoun to refer to an object during instructions. The effects that these reference types had on the time difference measure were investigated. The increased time taken for speakers in the no roles condition relative to the roles condition was found only for instructions using proper names but not pronouns. This finding is likely due to speakers using pronouns when objects and object locations had already been established, as this is typically when pronouns are used (Meyer & Bock, 1999; Schmitt, Meyer, & Levelt, 1999). Therefore knowing where the object was located before the procedure began had no effect on the time taken to fixate the object in these cases. The effect of pronouns on the time difference measure for listeners can be accounted for with this explanation. These participants were significantly quicker when pronouns were used, suggesting that the speaker used pronouns because they knew that the listener was aware of (or already looking at) the target location, much like Clark & Krych (2004) found that a speaker's utterances would change (and simplify and shorten) based on what the speaker knows the listener knows. When the speakers in the present study knew the listener was already looking, or had already previously attended to an object, they may have shortened their utterances accordingly (using pronouns), speeding up communication.

A similar analysis was carried out for verbal references delivered alongside or without verbal spatial information. The presence or absence of these cues was not found to have any effect on the time for the speaker to fixate a referent object. Listeners, however, were found to be much quicker to fixate the object in the roles condition when non-spatial references were used, compared to other instances. Like the use of pronouns, it may be the case that the speaker used non-spatial instructions when the listener already knew the location of the referent object. This would explain why listeners were much quicker to locate non-spatially referenced items in the roles condition, as the speaker may not be using spatial language as they know that the listener knows the spatial location of the target object. If both the use of pronouns and non-spatial language were related to the speaker's understanding of the listener's knowledge of the referent object location then these factors would be expected to cooccur. Fewer spatial references were found to be used alongside pronoun references compared to object name references, however this difference was only approaching significance.

The above results suggest that the use of spatial or non-spatial language in this task was influenced not only by what the speaker knew, but also by what the speaker knew that the listener knew. As with the findings for the use of pronouns, these findings were in line with previous research showing that speakers alter their utterances online, as their understanding of what the listener knows changes (Clark and Krych, 2004). It is unclear from the present study whether the changes in gaze allocation found across different language types were the result of the same factors that affected the use of different types of language or whether the use of different language types directly affected gaze allocation. However, in a more controlled follow up experiment (Chapter 3), the naturally occurring linguistic references that have been identified in this study will be manipulated to investigate the interaction between gaze utilisation and reference type.

This study investigated the effect of social roles on language and gaze behaviour in a natural collaboration using dual portable eye-trackers. In this real social collaborative setting, people spent very little time looking at each other, challenging the generalisability of the conclusions from lab-based paradigms (Birmingham et al, 2007; 2009; Zwickel and Võ, 2010). These results provide a strong case for investigating gaze cueing behaviour in highly naturalistic environments, which is how this thesis will continue to investigate the interaction between gaze cueing and language.

This chapter provides evidence to suggest that those in the roles condition were more closely attending to where their partners were looking. Listeners were found to look more at a speaker providing verbal instructions if the speaker was playing the role of a chef, suggesting that our tendency to look at others is either affected by our social perceptions of a person or by our perception of their reliability. These results provide evidence that social context can affect the utilisation of gaze cues.

Additionally, the present study has shown that in these collaborations the form and type of the spontaneous verbal references used varied between instructions, and as these varied gaze allocation varied. Spatial verbal references sometimes used object names and spatial language, or pronouns and non-spatial language. The present chapter speculated as to why these different types of reference were used and discussed the effects on visual behaviour, however there is a limit to what can be concluded from this largely uncontrolled naturalistic study. Following the cognitive ethology framework (Kingstone et al., 2008), having identified the spontaneous use of different reference types and their effect on eye movement behavior in natural interactions, this thesis will progress by exploring the effects of manipulating different types of verbal reference on gaze cue utilisation in a more controlled environment.

Chapter 3:

Referent expressions and gaze cues – Study 2

Introduction

The previous chapter outlined evidence that social context and the type of language we use affects how we allocate our gaze in social interactions. Participants in Study 1 varied the type and form of the verbal references used throughout the interactions, but it is not clear how changing the nature of the information conveyed in speech might influence whether and how gaze cues are utilised. Specifically, referring expressions can contain featural or spatial determiners that unambiguously specify the target object and examples of both types of determiner were found in the natural collaboration study outlined in the previous chapter. Given the inherently spatial nature of a gaze cue, a concurrent featural referring expression offers information additional to and non-overlapping with that conveyed by a gaze cue, whereas a concurrent spatial referring expression offers information that largely overlaps with the gaze cue. The present chapter explores whether and how gaze cue utilisation varies depending on whether concurrent spoken language refers to objects using spatial or featural disambiguating determiners.

In Macdonald and Tatler (2013) the informativeness of the language was varied by using referring expressions that either unambiguously identified an object or were ambiguous and equally described two objects in the array. In this way, gaze cues provided no additional information to that provided by unambiguous spoken instructions. However, gaze did provide additional and essential information for locating the target object when provided alongside ambiguous spoken instructions. Gaze cues therefore either provided entirely redundant information or were the only source of information that reliably signalled the location of the correct object. The aim of the present study was to investigate whether the type of information conveyed in language can vary gaze cue utilisation in the same way as varying the informativeness of language. Inspiration for how to vary language in this study came from the finding in Study 1 that instructions with spatial language were associated with increased time to fixate referent items. Here, the type of disambiguating determiner (featural or spatial) used in an instruction was manipulated. Gaze cues are inherently spatial cues; they direct attention to the position of an object, but are not capable of providing any information about the features of the object, such as colour or shape. Therefore, a gaze cue may be able to communicate the same information as a verbal spatial reference, but not a verbal featural reference, making gaze cues potentially more informative alongside featural references. The present study considered whether people's gaze seeking and following behaviour reflects the increased informativeness of a gaze cue provided with a featural descriptor compared to a spatial descriptor.

Despite the fact that studies of gaze cuing in more natural contexts appear to argue against reflexive gaze following, there is previous evidence that automatic and selective gaze utilisation may occur together. Laidlaw, Risko and Kingstone (2012) presented participants with static images of faces and instructed participants to not look at either the mouths or eyes. Compared to free viewing, participants managed to easily avoid fixating mouths, but struggled with eyes. However, when carrying out this task on pictures of inverted faces, participants could equally suppress looks to mouths and eyes. Combined, these results suggest orienting to another's eyes is both selective and automatic and that the automatic effect may be linked to holistic face processing. Hill et al. (2010) found evidence of automatic gaze following in a Posner-type (1980) task when targets were presented less than 200 ms after gaze cue onset, but selective gaze following when targets were presented more than 200 ms after gaze cue onset. These studies provide good evidence for the existence of both automatic and selective components of gaze cue utilisation; however it is still not clear whether the automatic component has any effect in a real world social interaction, in which a number of complex factors influence the use of gaze cues.

In these previous studies, key evidence suggesting an automatic component of gaze following has been provided by situations in which participants follow gaze cues that do not reliably signal information about the target (such as gaze cues that cue a direction incongruent with the target). Therefore, this paradigm includes an incongruent gaze condition, to allow automatic attentional effects to be inferred from any disruption in task performance caused by incongruent gaze cues.

Study 1 found that our social perceptions of a partner in a task could affect how we allocate gaze to and with that partner. There is also evidence the social skills of an individual can affect gaze orienting and following behaviour. Laidlaw et al. (2011) found the extent to which participants orient to a present stranger was inversely correlated with a measure of the participants' autistic traits. Freeth, Foulsham & Kingstone (2013) found the converse effect when people were actually involved in an interaction; high levels of autistic traits were associated with decreased looking time at a person speaking in a video, but not in real one-to-one interactions. In the present study, participants completed the same Autism-Spectrum Quotient (AQ) questionnaire (Baron-Cohen, Wheelwright, Skinner, Martin & Clubley, 2001) used by Laidlaw et al. (2011) and Freeth et al. (2013) to investigate any effect autistic traits may have on the eye movement measures.

In the present study, each participant followed instructions to complete a search task, while wearing a portable eye-tracker. The instructor varied the verbal determiner used at the end of the instruction (featural or spatial) and the presence of gaze cues (absent, congruent or incongruent). The present study aims to first investigate whether the type of verbal reference used influenced the extent to which gaze cues were sought out and followed. Secondly, the incongruent gaze condition was used to investigate whether a real-world complex task can be influenced by the low level attentional shifts caused by gaze cues. We hypothesised that participants would utilise the instructor's gaze flexibly, based on the informativeness of the gaze cue relative to the language used in the instruction.

Method

Participants

Twenty-four undergraduate students (female = 23) from the University of Dundee were recruited for this study. They received course credit for participation. They were split equally into three groups: one for each of the gaze conditions.

Materials

Forty experimental items and 20 distracter items were used for this experiment. The 40 experimental items were made up of 20 object pairs. The objects in each pair were identical, except for their colour (see Figure 3.1). All of the items were used in each of the eight sets of trials (each set involved five trials). For each set of trials one of eight different object layouts was used. Each layout involved the objects positioned equally across four surfaces (two on table tops, two under tables). To counterbalance for any directional biases, half of the participants used an alternative set of layouts, which were the mirror image of the initial eight layouts. The instructor used a unique instruction sheet for each participant. Each trial was counter-balanced for the location, colour and type of determiner used for the target across participants.

Figure 3.1. An example of one of the 20 object pairs. These purses differ only in colour.

Each participant filled out an Autism-Spectrum Quotient (AQ) Questionnaire (*Appendix 6)*, devised by Baron-Cohen et al. (2001). The questionnaire involved 50 statements that were designed to measure the degree of self-reported autistic traits in non-clinical populations. Participants had the same four multiple-choice responses to each statement: “Definitely agree”, “Slightly agree”, “Slightly Disagree” and “Definitely disagree”. Each questionnaire was scored out of 50 according to the guidelines set by Baron-Cohen et al. (2001) (See *Analysis of Autism-Spectrum Quotient Questionnaire* section).

Design

This experiment used a 2 (within subjects) × 3 (between subjects) design. The independent variables were type of determiner used (featural or spatial) and gaze condition (absent, congruent or incongruent). A between-subjects design was used for gaze condition to give participants the opportunity to learn the type of cue to expect from the instructor, so that potential gaze seeking strategies could be investigated. In order for participants to learn instructor gaze behaviour, trials had to be blocked by gaze condition. In a within-subjects design, this would require a large number of trials per participant, making analysis overly laborious. Furthermore, in a within-subject design learned behaviours from one block may influence another. Therefore only one gaze condition was assigned to each participant.

Procedure

In addition to the mobile eye tracker worn by the participant throughout each recording session, two extra cameras were used. One was placed so that it would capture the instructor (experimenter) throughout the procedure and the other was placed so that it would capture the participant. The eye tracker was calibrated (see *Eye Movement Recording* section). The participant was positioned in front of a bag facing the instructor, who was approximately four metres away (see Figure 3.2). The participant was aware that the instructor was an experimenter, but was not made aware that he was investigating gaze cueing. Between the participant and instructor and on either side of them were two tables. There were 15 objects on each table-top and 15 objects underneath each table, but still visible from the participant's and experimenter's starting positions. Of the 60 objects, 40 were experimental objects and five of these were the targets for the first set of five trials. The counterpart (identical object with different colour) to each of these five target objects was positioned on the opposite side to the target. When experimental items were not targets for any given layout, objects were sometimes positioned on the same side as their counterpart (as is the case with the green and yellow frisbees in Figure 3.2).

Figure 3.2. A still image from the synchronised digital video recordings. The panel on the left shows the instructor providing a gaze cue and the panel on the right shows the participant standing opposite the instructor. The central panel shows the scene and right eye recordings from the eye-tracker. The cross-hair shows the participant's point of fixation.

The instructor explained to the participant that they would be told to pick up objects, one at a time. After each instruction they were required to find the object, pick it up and then put it in the bag, before returning to their starting position. For each instruction the instructor looked at the participant before beginning. Each instruction statement followed the same structure; they began with the words “Pick-up the [object name]” and were followed by either “on the [left or right]” (spatial condition) or “that’s [object colour]” (featural condition). Before the experiment began, the experimenter explained that spatial references would refer to locations with respect to the participant rather than with respect to the experimenter giving the instruction. Thus an instruction to “pick up the purse on the right” would mean the purse to the right of the participant. After five instructions were given the participant was escorted behind a partition and the objects were re-arranged into a pre-determined layout. Eight layouts were used in total for each participant and the order was randomised. Of the 40 instructions, 20 used a spatial determiner (left or right) and 20 used a featural determiner (object colour). In the congruent gaze cue condition, the instructor looked at the target object after stating the object name, but before giving the determiner. At the same point in the incongruent condition, the instructor looked at the counterpart to the target object, on the other side of the room. In the no gaze condition the instructor looked at neither object, but instead looked down to the instruction sheet. After the main procedure participants completed the Autism-Spectrum Quotient (AQ) questionnaire.

Eye Movement Recording

Participants' eye movements were tracked using a Positive Science LLC mobile eye tracker (New York, NY). This was one of the eye-trackers used in Study 1, and videos were recorded and rendered and eye position estimated using the methods outlined in study one. Calibration involved asking the participant to look at particular points on a pin board at a distance of about three metres, followed by particular toy building blocks (Megabloks) on a counter top at a distance of 70 cm. Calibration was repeated after the task. Sound was also recorded throughout the experiment.

Analysis of eye-movement and audio data

Eye-tracking data were manually coded offline by the first author and a volunteer research assistant. There was no minimum fixation duration criterion. The first author coded the timings of looks to the target and counterpart objects as well as the timings of gaze cues. The first author and a research assistant coded the number of looks to the instructor and length of looks to the instructor for each trial. The lead experimenter and research assistant initially coded the same movie file and these were compared by the lead experimenter to ensure a consistent and high quality of coding. Looks to the instructor were considered to be any fixations that were on any part of the instructor's body. This liberal criterion, also used in Macdonald & Tatler (2013), was employed so that any looks to the instructor were classed as potential instances of foveal or parafoveal gaze-seeking behaviour.

Audacity sound editing software was used to extract the timings for the onset of each instruction statement, the onset of each object name and the onset of each determiner word. Of the 960 experimental trials from the 24 participants, 144 were discarded, due to three of the objects being commonly misidentified, leaving 816 potentially usable trials. There was further data loss, due to camera errors or loss of pupil/corneal reflection tracking, which affected different dependent variables in different ways. The number of trials used in the analysis of each dependent variable is noted below.

The first three dependent variables (DVs) were indicators of gaze seeking and focussed on looks to the instructor. These were: (1) proportion of trials in which the participants fixated on the instructor at trial onset (779 analysed trials), (2) proportion of trials in which the participants fixated on instructor at the time of gaze cue (744 analysed trials) and (3) the mean proportion of time in each trial the participant spent fixating the instructor (642 analysed trials). The instructor provided gaze cues after the onset of the object name, but before the onset of the disambiguating word. Gaze cue onset times were identified from the video footage of the instructor (left-most panel in Figure 3.2). On average, the gaze cue was provided 791ms after the onset of the object name. Since no gaze cues were given in the no gaze cue condition, that time point could not be used. Instead, the mean time the gaze cue was given in the incongruent and congruent conditions relative to the onset of the object name and onset of the determiner was calculated. On average, gaze cues were given after .88 of the time between the onset of the determiner word and object name had elapsed. For each trial in the no gaze condition, .88 of the time difference between the onset of the determiner word and onset of the object name was calculated. This was added to the time of onset of the object name to calculate the "time of gaze cue". When calculating the proportion of time spent looking at the instructor, the onset of the instruction sentence was used as the beginning of the trial and the pick-up time was used as the end of the trial.

The remaining two dependent variables were used as indicators of gaze following. Firstly, (4) the proportion of trials in which participants initially looked in the correct direction after the onset of the gaze cue (750 analysed trials). This allowed for the investigation of initial responses to congruent and incongruent gaze cues. The effect of the conditions on performance in the task was investigated by measuring (5) the time difference between the first look at the target and the onset of the trial (705 analysed trials). Smaller time differences between the first look at the target object and the onset of trial were used as indicators of better performance.

Time-course of trials

Given that the stimuli used in this experiment were produced and presented by a present experimenter in real time, there is naturally some variation in the timings of the language and gaze cue onsets. On average the onset of the object name occurred 428 ms after the beginning of the instruction sentence. When a gaze cue was provided, this occurred a mean 791 ms after the onset of the object name. The onset of the disambiguating word occurred on average, 96ms after the gaze cue, 887 ms after the onset of the object name and 1313 ms after the beginning of the senctence. Mean total trial time (time between onset of sentence and a participant touching an object) was 5528 ms.

Analysis of Autism-Spectrum Quotient Questionnaires

The responses to each statement were scored according to the guidelines set out by Baron-Cohen et al. (2001). Half of the statements were congruent with an autistic trait (e.g. "I would rather go to a library than a party") and the remainder of the statements were incongruent with an autistic trait (e.g. "I find social situations easy"). A point was given to each "Definitely agree" or "Slightly agree" response to autistic-trait-congruent statements and to each "Slightly Disagree" and "Definitely disagree" response to autistic-trait-incongruent statements. The total point-score out of 50 was considered the participant's AQ score. The more the participant identified with autistic traits, the higher the score. Baron-Cohen et al. (2001) found that a subject pool of adults with high-functioning autism or Asperger syndrome had a mean score of 35.8 (SD = 6.5) and a cohort of control subjects scored a mean of 16.4 (SD = 6.3). The scores in the present study ranged from 7-24, with a mean of 16.8 (SD = 4.7), showing results typical of a normal population. The participants were divided into two groups: a low AQ group of the eleven participants that scored between 7-17 (Mean = 13.3, SD = 2.2) and a high AQ group of the 13 participants that scored between 18-24 (Mean = 19.8, SD = 4.2). Of the eleven participants in the low AQ group, six were in the no gaze condition, two were in the congruent condition and three were in the incongruent condition. Of the 13 participants in the high AQ group, two were in the no gaze condition, six were in the congruent condition and five were in the incongruent condition.

Statistical analysis

The data were all analysed using the *lme4* package (Bates, Maechler, Bolker & Walker, 2014) in the R statistical programming environment to run linear mixed effects models (LMMs). As with Chapter 2, the details of the approach are outlined in this section.

Random Factors. The majority of LMMs in this chapter used subject and item as random factors. However, the models investigating the proportion of trials in which the instructor was fixated at the onset of the trial do not include item as a random factor. This is due to this measure occurring early in the trial, before item could theoretically have any effect on performance. Random slope models were used for the reasons outlined in Chapter 2. Again *maximal* models were used where possible.

Simplifying models. When these maximal models failed to converge and required simplification of the random effects structure, the approach was to first remove the correlations between random-slopes and intercepts. After this, the slope of any interaction from the random factor was removed, starting with the item random effect (if present). If necessary, further simplification was made by selectively removing the slopes of the interaction and main effect terms, initially in the item-level random effect structure (see *Appendix 1* for more details).

Sum coding, calculating *p*-values and pairwise comparisons. These procedures were all carried out as outlined in Chapter 2.

Results

Gaze seeking: Looks to the instructor

The first measure was concerned with any effect the between-subjects gaze conditions had on whether or not participants fixated the instructor at the onset of each trial. The mean proportion of trials in which this occurred for each gaze condition is shown in Figure 3.3a. Participants were nearly always looking at the instructor when the trial began in both the no gaze (.97) and congruent gaze (.98) conditions, but fewer trials began with a look to the instructor in the incongruent gaze condition (.76). An LMM (Model 3.1) using gaze condition as a fixed factor (and subject as a random factor) confirmed that there was no significant difference between the congruent and no gaze conditions, β = .149, *SE* = 1.303, z = .115, p = .909, but there was a significant difference between the no gaze and incongruent conditions, β = -2.844, *SE* = 1.170, z = -2.431, p = .015.

To investigate if AQ had any influence on whether participants in the incongruent condition looked at the instructor, An LMM (model 3.2) using AQ group as a fixed factor and subject as a random factor was carried out. There was no significant difference in the mean proportion of trials in which the instructor was fixated at the onset of the sentence between the low AQ group (.74) and high AQ group (.76), β = 0.1369, *SE* = 1.702, z = 0.080, p = . 936.

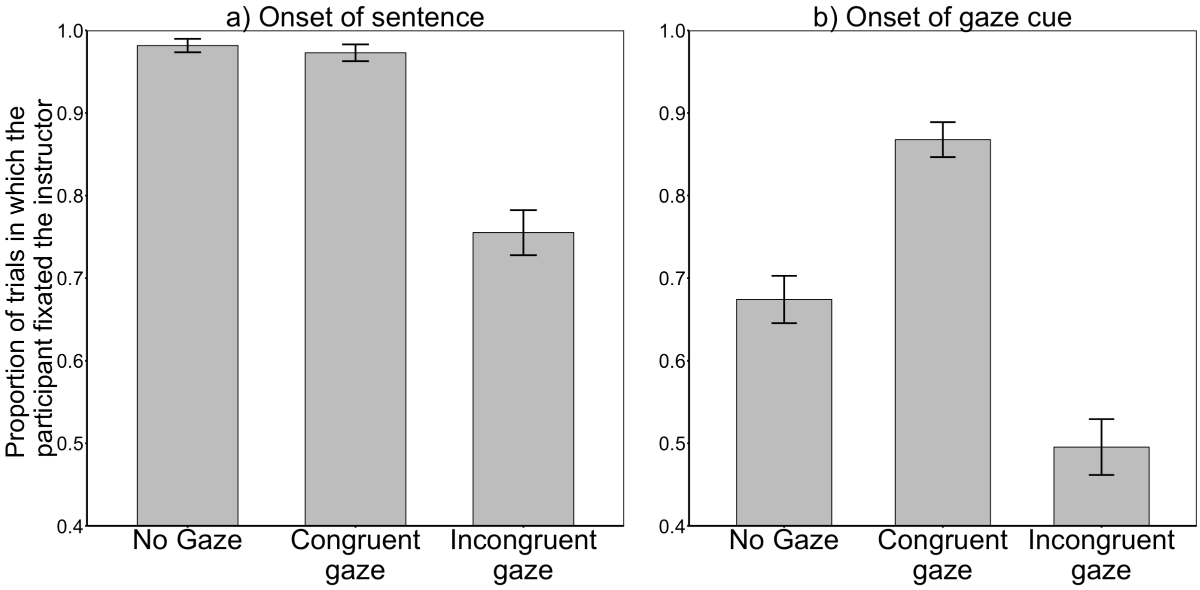

Figure 3.3. The proportion of trials in which the participant fixated the instructor a) at the onset of the instruction sentence (beginning of trial) and b) at the onset of the gaze cue. Results are shown (with standard error) for the No gaze, Congruent gaze and Incongruent gaze conditions.

The proportion of trials in which the participants fixated the instructor at the onset of the gaze cue was measured (Figure 3.3b). In most trials, at the onset of the gaze cue participants in the congruent gaze condition were looking at the instructor (.87). The proportion of looks to the instructor at this point was lower for the no gaze condition (.67) and lower still for the incongruent condition (.50). An LMM (Model 3.3) using gaze condition as a fixed factor (and subject and item as random factors) showed the difference between the no gaze condition and the congruent condition to be non-significant, β =1.533, *SE* = 1.039, *z* = 1.476, *p* = .140. The difference between the incongruent and no gaze conditions was also non-significant, β =-1.008, *SE* = 1.029, *z* = -.980, *p* = .327, but the difference between incongruent and congruent conditions was significant, β =-2.541, *SE* = 1.047, *z* = -2.427, *p* = .015.

As with looks at the onset of the sentence, AQ analysis was focussed on whether AQ influenced the extent to which participants fixated the instructor at the onset of the gaze cue. As none of the gaze conditions approached 100% trials with looks to the instructor, all conditions were investigated using LMMs with AQ group as the fixed factor and subject and item as random factors. There was no significant difference found between AQ groups for the congruent condition (Model 3.4), β =-1.266, SE = 2.063, z = -.614, p = .540, the incongruent gaze condition (Model 3.5), β =-1.154, SE = 1.807, z = -.639, p = .523, nor in the no gaze condition (Model 3.6), β =-2.048, SE = 1.497, z = -1.368, p = .171.

The mean proportion of trial time spent looking at the instructor for each gaze and reference condition is shown in Figure 3.4. An LMM (Model 3.7) for the congruent and no gaze conditions (using subject and item as random factors) showed no significant effect of gaze condition, β =.039, SE =.028, t = 1.378, p = .155, but a significant effect of reference condition, β = -.017, SE =.008, t = -2.066, p = .043, and a significant interaction, β = -.033, SE =.016, t = -1.994, p = .049. Pairwise comparisons showed that the mean proportion of trial time that the participant spent looking at the instructor was significantly greater for featural trials than spatial trials in the congruent condition, p = .011 (mean featural trial time = 5,126 ms, mean spatial trial time = 5,325 ms), but not the no gaze condition, p = .999 011 (mean featural trial time = 5,380 ms, mean spatial trial time = 5,478 ms).

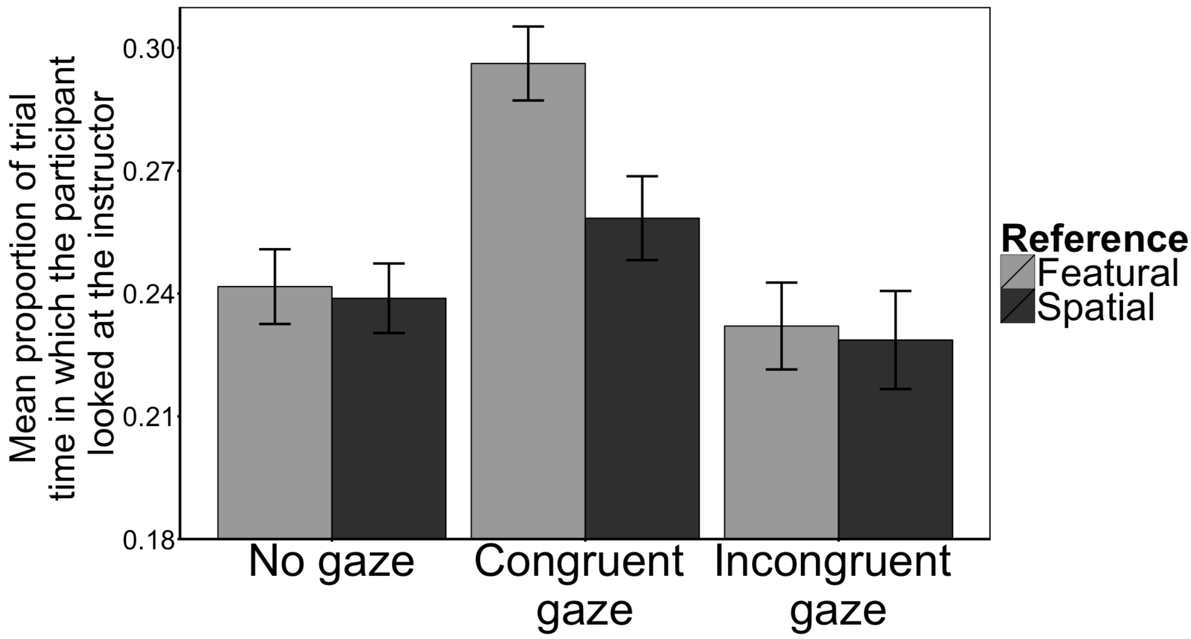

Figure 3.4. The proportion of trial time that the participant spent looking at the instructor. Results are shown (with standard error) for the No gaze, Congruent gaze and Incongruent gaze conditions for both featural and spatial reference conditions. Mean trial time throughout the experiment was 5,528 ms.

An LMM (Model 3.8) for the no gaze and incongruent gaze conditions (using subject and item as random factors) showed no effect of gaze condition, β =-.024, *SE* =.034, *t* = -.695, *p* = .466, nor reference condition, β =-.006, *SE* =.008, *t* = -.710, *p* = .478, nor any significant interaction, β =-.011, *SE* =.016, *t* = -.718, *p* = .482.

The congruent gaze data was analysed further to investigate whether AQ influenced the extent to which participants looked at the instructor and whether any effect interacted with reference type (Figure 3.5). The LMM (Model 3.9) used reference type and AQ group as fixed factors and subject and item as random factors. A significant effect of reference type was found with this model, supporting the findings of the previous Tukey test, *β* =-.035, *SE* =.013, *t* = -2.579, *p* = .014. The model also showed a significant effect of AQ group, *β* =-.075, *SE* =.026, *t* = 2.902, *p* = .009, but no significant interaction, *β* =-.038, *SE* =.034, *t* = 1.127, *p* = .252. Despite the lack of a significant interaction in the model, there was *a priori* interest in whether the change in reference condition affected the AQ groups differently. Post-hoc Tukey tests revealed that the difference between featural and spatial reference instructions was significant in the low AQ group (*p* = .038), but not the high AQ group (*p* = .355).

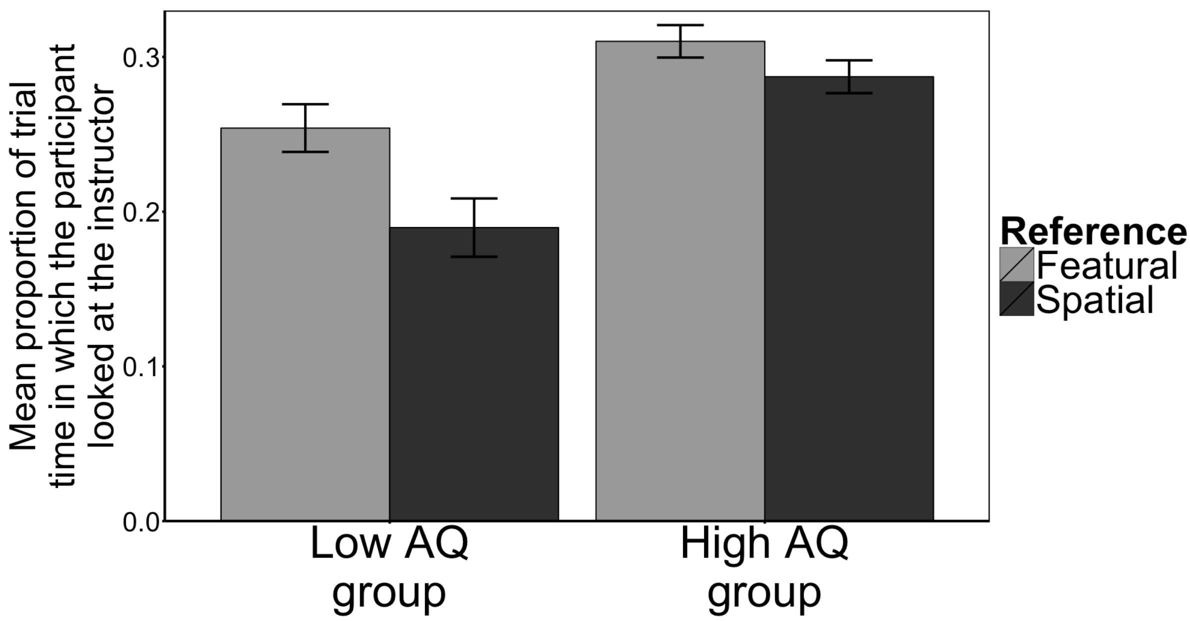

Figure 3.5. The proportion of trial time that the participant spent looking at the instructor for the congruent gaze condition only. Results are shown (with standard error) for the high AQ group and low AQ group for both featural and spatial reference conditions. Mean trial time throughout the experiment was 5,528 ms.

The mean proportion of trial time that the participants spent fixating the instructor did not significantly vary across the no gaze and incongruent gaze conditions, nor did it vary across the featural and spatial reference conditions. Therefore, the influence of AQ group on all of the data from the no gaze and incongruent conditions was analysed together. Despite a difference between the low AQ (.25) and high AQ (.21) groups, an LMM (Model 3.10) using AQ group as a fixed factor and subject and item as random factors showed no significant difference in the mean proportion of trial time spent fixating the instructor between the low and high AQ groups, β =-.046, *SE* =.033, t = -1.403, p = .148.

The results from the analysis of looks to the instructor suggest that participants engaged in selective gaze seeking behaviour. Participants looked to the instructor at the onset of the gaze cue more frequently when these cues were anticipated to be congruent and they spent more time looking at the instructor overall when the gaze cue provided the most uniquely informative information (when presented alongside featural language in the congruent gaze condition).

Initial gaze following

Initial gaze following was investigated by measuring the mean proportion of trials in which the first saccade launched after the gaze cue was in the direction (left or right) of the target object (Figure 3.6a). The first eye movement was usually in the correct direction in the congruent gaze condition for both spatial (.92) and featural (.89) references. There were fewer trials with correct first initial saccades in the no gaze/spatial condition (.79) and fewer still in the no gaze/featural condition (.56). An LMM (Model 3.11) of the congruent and no gaze conditions (using subject and item as random factors) showed a significant effect of gaze condition, β =.1.532, *SE* =.296, z = 5.185, p < .001, and reference type, β =.687, *SE* =.259, z = 2.653, p = .008, but no significant interaction, β =-.775, *SE* =.523, z = -1.482, p = .138. Pairwise comparisons showed a significant difference between the congruent and no gaze conditions in featural reference trials, p < .001, and in spatial reference trials, p = .028.

From Figure 3.6a it appears that there was little difference in the results for the incongruent gaze condition and no gaze condition for both spatial (.74) and featural (.56) references. An LMM (Model 3.12) of the incongruent and no gaze conditions (using subject and item as random factors) confirmed that there was no effect of gaze condition, β =-.125, SE =.231, z = -.541, p = .589, nor any interaction between gaze and reference conditions, β =-.247, SE =.399, z = -.618, p = .536. There was, however, a significant effect of reference type, β =.943, SE =.202, z = 4.678, p < .001.

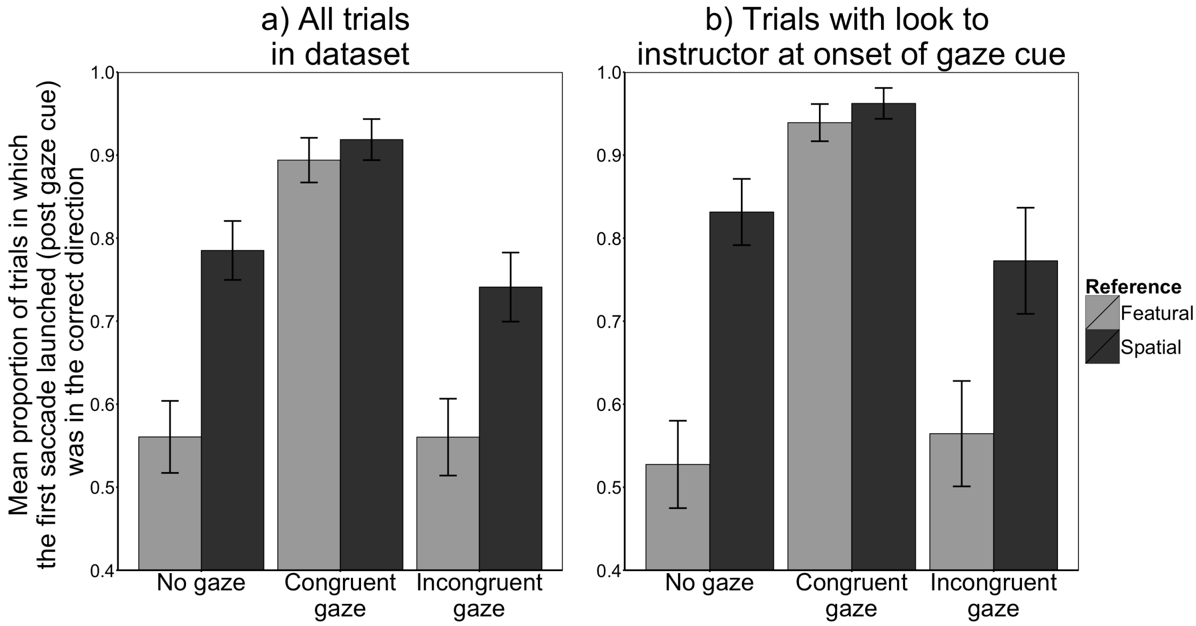

Figure 3.6. The mean proportion of trials with initial eye movements after the gaze cue in the correct direction for a) the full dataset and b) only trials in which the participant was looking at the instructor when the gaze cue was given. Results (with standard error) are shown for the No gaze, Congruent gaze and Incongruent gaze conditions for both featural and spatial reference conditions.

It is possible that performance in this measure may have been mediated by whether or not the participant was fixating the instructor at the point in the time at which the gaze cue was given. To investigate this, the analysis was repeated using only the trials in which the instructor was fixated at the onset of the gaze cue. This analysis included 87% of congruent trials, 67% of no gaze trials and 50% of incongruent trials (See Figure 3.3b). Results from this analysis are shown in Figure 3.6b.

The focus of interest in this re-analysis was to investigate whether the relationship between the results for those in the three gaze conditions changed, when analysing only those trials with fixations on the instructor at the onset of the gaze cue. The

LMM (Model 3.13) of congruent and no gaze trials showed the same effects as the LMM on the full dataset; there were main effects of gaze condition, β =2.154, *SE* =.364, z = 5.894, p < .001, reference condition, β = .942, *SE* =.389, z = 2.424, p = .015, but no significant interaction, β =-.984, *SE* =.732, z = -1.343, p = .179. From Figure 3.7, there appears to be little difference in the mean proportion of trials with initial eye movements in the correct direction between the no gaze (spatial = .83, featural = .53) and incongruent gaze (spatial = .77, featural = .56) conditions. The LMM (Model 3.14) showed that there was no significant main effect of gaze condition, β =-.149, *SE* =.456, z = -.327, p = .744, nor an interaction, β =-.572, *SE* =.657, z = -.871, p = .384, but there was a significant main effect of reference type, β =1.364, *SE* =.317, z = 4.300, p < .001; this replicates the pattern of effects found in the analysis of the full dataset.

A similar analysis using only the trials in which the instructor was not fixated at the onset of the gaze cue would not be appropriate given the small percentage of trials in which this occurred in the congruent (13%) and no gaze (33%) conditions. However, for the incongruent gaze condition only, an LMM (Model 3.15) was used to investigate any interaction between reference type and whether the participant looked at the instructor at the onset of the gaze cue (Figure 3.7).

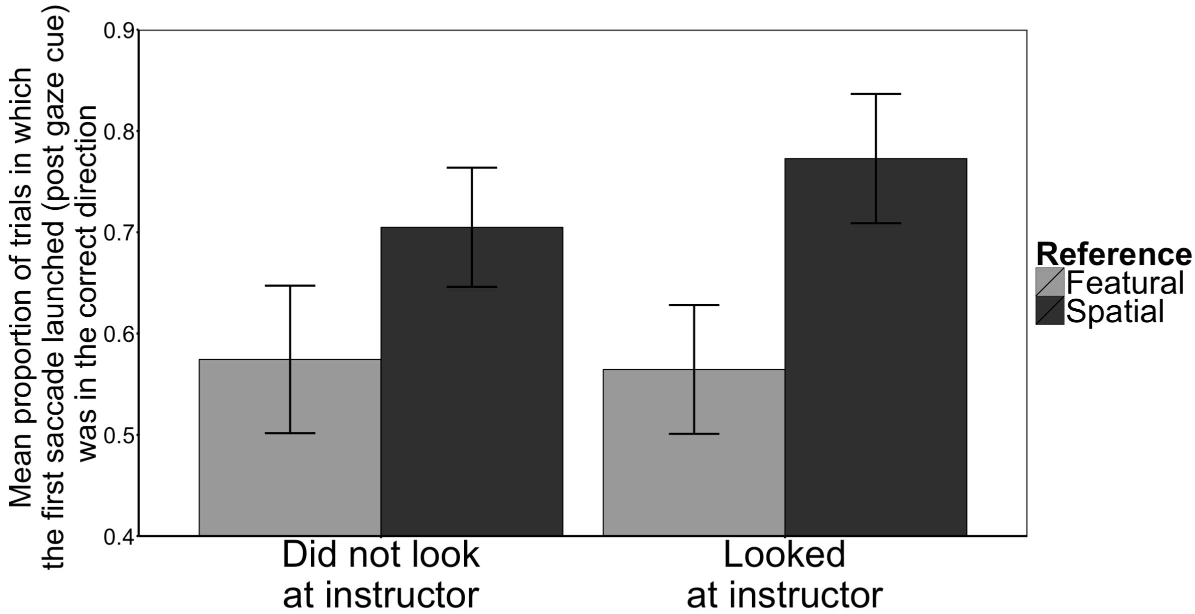

Figure 3.7. The mean proportion of trials with initial eye movements after the gaze cue in the correct direction. Results are displayed for the Incongruent gaze condition only and separated by reference condition and by whether or not the instructor was fixated at the onset of the gaze cue.

The LMM showed there to be no difference in performance between the trials in which the instructor was fixated and those in which he was not, β =.137, *SE* =.478, z = .286, p = .775. A significant effect of reference type was found, β =.892, *SE* =.351, z = 2.543, p = .011, but there was no interaction, β =.463, *SE* =.656, z = .705, p = .481.

As with the measure of proportion of trial time spent fixating the instructor, the influence of AQ group on the proportion of first saccades in the target direction was investigated. LMMs using AQ group as a fixed factor and subject and item as random factors showed that there was no significant effect of AQ group in the congruent condition (Model 3.16), β =-.764, *SE* =1.100, z = -.694, p = .487, nor the incongruent and no gaze conditions (Model 3.17), β =-.286, *SE* =.217, z = -1.319, p = .187.

The analysis of first saccade direction after the gaze cue onset showed that saccades were no more likely to be launched in the wrong direction following an incongruent gaze cue than in the absence of gaze cues (Figure 3.6). This is somewhat in contrast to the increased proportion of erroneous gaze-following responses to incongruent gaze cues typically found in Posner type gaze-cueing paradigms (Ricciardelli et al., 2002; Kuhn & Kingstone, 2009). However, it may be that any distracting effect of the incongruent gaze cue might manifest not as differences in overt responses but as differences in covert attention allocation. Such effects might result in a delayed overt response following an incongruent gaze cue rather than an incorrectly-directed overt response. The next analysis explores the possibility of such covert effects of gaze cues by comparing the time taken to first fixate the target, which could have potentially been influenced by attentional effects that did not result in the launching of saccades in the wrong direction.

Gaze following: Looks to objects

The effect of the conditions on performance in this task was investigated by measuring the mean time difference between the onset of the instruction sentence and the first fixation on the target object (Figure 3.8a). An LMM (Model 3.18) of the congruent (spatial = 70 frs., featural = 67 frs.) and no gaze conditions (spatial = 76 frs., featural = 86 frs.), with subject and item as random factors showed a significant effect of gaze condition, $\beta = 12.550$, $SE = 3.857$, $t = 3.254$, $p = .003$, and a significant interaction, $\beta = -12.625$, $SE = 6.173$, $t = -2.045$, $p = .040$. Reference type did not have a significant effect, $\beta = -3.577$, $SE = 3.006$, $t = -1.190$, $p = .212$. Pairwise comparisons showed a significant difference between the no gaze and congruent conditions for instructions with featural references, $p < .001$, but not spatial references, $p = .504$.

The LMM (Model 3.19) of the incongruent (spatial = 84 frs., featural = 83 frs.) and no gaze conditions, with subject and item as random factors, showed no significant effect of gaze condition, $\beta = -3.169$, $SE = 5.494$, $t = -.577$, $p = .546$, reference type, $\beta = -4.480$, $SE = 3.121$, $t = -1.435$, $p = .150$, nor any interaction, $\beta = 10.944$, $SE = 6.850$, $t = -1.598$, $p = .111$. As there was *a priori* interest in the different effects of reference type in different gaze conditions, pairwise comparisons were carried out despite the lack of a significant interaction. The comparisons showed no significant difference between the featural and spatial instructions for those in the incongruent condition, $p = .993$, and an approaching significant difference for those in the no gaze condition, $p = .061$.

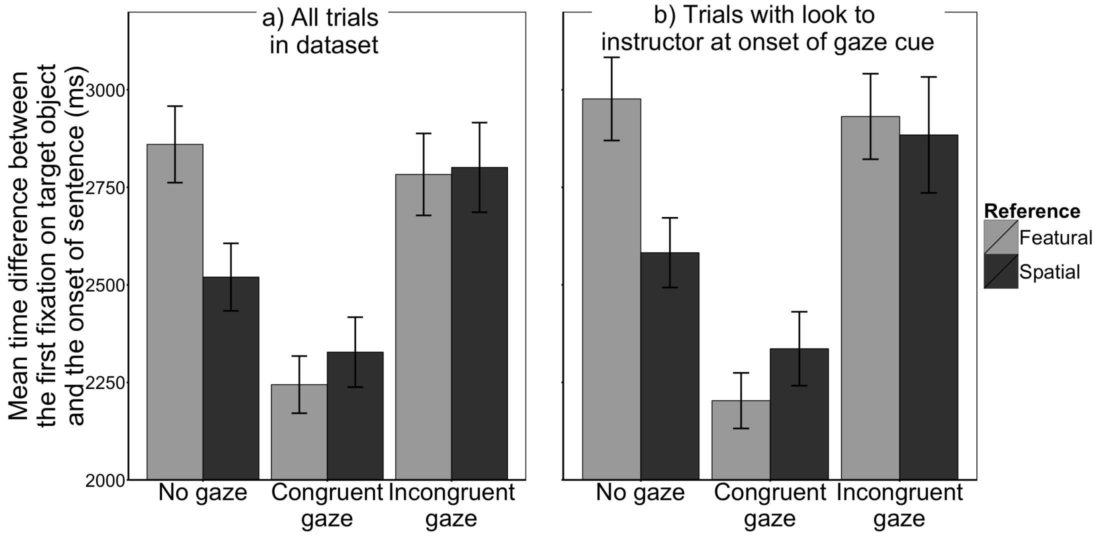

Figure 3.8. The mean time difference (frames) between the onset of instruction sentence and the first fixation of the target object for a) the full dataset and b) only trials in which the participant was looking at the instructor when the gaze cue was given. Results (with standard error) are shown for the No gaze, Congruent gaze and Incongruent gaze conditions for both featural and spatial reference conditions.

As with the previous measure, it is possible that the time difference between the onset of the sentence and the first fixation on the target object may have been affected by whether or not the participant was looking at the instructor when the gaze cue was provided. Therefore analysis was repeated for only those trials in which the instructor was fixated at the onset of the gaze cue (Figure 3.8b). The pattern of results in Figure 3.8b is very similar to the pattern in Figure 3.8a. The LMM (Model 3.20) for congruent and no gaze conditions showed a significant effect of gaze condition, β = -15.397, SE = 2.676, t = -5.755, p < .001, and an interaction, β = 15.873, SE = 6.428, t = 2.469, p = .015, but no effect of reference type, β = -2.76, SE = 3.045, t = -.919, p < .328. The LMM (Model 3.21) of the incongruent and no gaze conditions showed an approaching significant effect of reference type, β = -7.178, *SE* = 4.182, *t* = -1.720, *p* = .081, but no significant effect of gaze condition, β = 4.058, *SE* = 4.351, *t* = .930, *p* = .331, nor any interaction, β = 12.435, *SE* = 8.434, *t* = 1.470, *p* = .132. Because there was *a priori* interest in how verbal reference type may mediate gaze following, planned comparisons were performed despite the lack of a significant interaction. No significant difference was found between the featural and spatial instructions for those in the incongruent condition, *p* = .999, but there was a significant difference for those in the no gaze condition, *p* = .016.

For the incongruent condition only, an LMM (Model 3.22) was carried out on the time difference between the onset of sentence and the first fixation on the target object, using whether the instructor was fixated at the onset of the gaze cue and reference type as fixed factors. Participants in the incongruent condition took longer to make their first fixation on the target object when they looked at the instructor as he provided the gaze cue (spatial = 88.6 frs., featural = 90.7 frs.) compared to when they did not fixate the instructor (spatial = 77.0 frs., featural = 73.3 frs.). This difference was not significant, β = 8.633, *SE* = 6.763, *t* = 1.276, *p* = .180. This finding is limited, however, in what it can tell us about the effect of incongruent gaze cues. The time difference here may simply be a result of the participants who are not looking at the instructor at the onset of the gaze cue, having a head-start on the search, as they begin looking at the tables earlier.

The influence of AQ group on the time to first fixate the target was investigated for each gaze condition. LMMs using AQ group as a fixed factor and subject and item as random factors showed the there was no significant effect of AQ group in the congruent condition (Model 3.23), β = 6.187, *SE* = 5.252, *t* = 1.178, *p* = .229, the incongruent condition (Model 3.24), β = .987, *SE* = 9.110, *t* = .108, *p* = .897, and no gaze conditions (Model 3.25), β = -6.620, *SE* = 8.107, *t* = -.817, *p* = .371.

Discussion

In the present study a real world search task paradigm was used to investigate the effect of the type of verbal reference used in an instruction on gaze cue utilisation. The real world paradigm used real gaze cues and spoken language and measured indicators of gaze seeking and following. The results showed clear signs of selective gaze utilisation behaviour, but also some indication that gaze cues affected the time to fixate the target, even when the cues were known to be unreliable.

Measuring the participants' looks to the instructor allowed for the investigation of gaze-seeking behaviour. The first of these measures was the proportion of trials in which the participant fixated the instructor at the start of the trial. Participants in the congruent and no gaze conditions were found to fixate the instructor at the beginning of most trials. However, there were significantly fewer trials with initial fixations on the instructor in the incongruent gaze condition. This is likely due to these participants avoiding what they anticipated to be unhelpful gaze cues. This shows strategic gaze seeking behaviour, however it may also suggest an automatic capture of gaze cue stimuli. Participants began each trial standing facing the instructor, so in order to begin a trial by not looking at the instructor, participants would have had to look away. If gaze cues were known to be unhelpful and gaze following was a completely selective process, then there would be no need to avoid these stimuli.

Therefore, the finding could suggest the participants were looking away because they anticipated a cue that would hinder them in the task. This explanation is in line with the typical findings of the gaze cueing paradigm (Friesen & Kingstone, 1998; Ricciardelli et al, 2002) in which known unhelpful gaze cues disrupt task performance, however these results clearly show that participants strategically altered their gaze seeking behaviour due to their expectation of gaze cue reliability.

The second gaze seeking measure was the proportion of trials in which the participant fixated the instructor later on in the sentence; specifically, at the point that the gaze cue was provided. For the no gaze condition, the average time that the gaze cue occurred in the other two conditions was calculated and was used to calculate a pseudo-gaze cue time for comparison (see Method). Participants looked at the instructor at the onset of the gaze cue in more trials in the congruent gaze condition compared to the incongruent gaze condition, showing that participants looked to the instructor more when they expected a helpful cue compared to when they expected an unhelpful cue. This supports the earlier findings (Macdonald & Tatler, 2013) that gaze cues were sought out more when they were more informative. It is unclear whether this difference is driven by those in the congruent condition actively seeking gaze cues, those in the incongruent condition actively avoiding gaze cues, or a mixture of these two behaviours.

The final gaze orienting measure was the mean proportion of trial time that participants spent looking at the instructor. This was used as an indicator of gaze seeking for featural and spatial determiner trials for participants in all three gaze conditions. In the previous chapter, the possibility was discussed that in a natural interaction long periods of viewing a collaborator may not be the ideal way to view a gaze cue. However, in this paradigm, gaze cues were maintained throughout the trial, meaning that all looks to the instructor after gaze cue onset could have potentially gathered information from gaze. Participants provided with featural references in the congruent gaze condition spent the most time looking at the instructor, suggesting that gaze cues were sought out the most when featural references were provided alongside an informative gaze cue. This condition was the condition in which the gaze cue was the most useful, relative to the spoken language being used. This is because the gaze cue was a) directed to the correct object and b) a unique source of spatial information. In the condition where congruent gaze cues were provided with a spatial determiner the gaze cue still reliably signalled the location of the target, but provided information that was also provided by language, making it no longer uniquely informative in providing the spatial location of the target object. The finding that gaze cues were oriented to for longer when they provided information not available from the spoken language provides further evidence for selective gaze seeking behaviour in real world interactions. Crucially, although gaze behaviour was predictable, reference condition was not, therefore participants must have selectively modified gaze-seeking behaviour during the trial, once the disambiguating word was processed and identified.

Combined, the measures investigating looks to the instructor provided evidence that participants adopted strategic gaze seeking behaviours, with those in the congruent gaze condition more likely to fixate the instructor at the onset of the gaze cue than those in the other conditions. Furthermore, within the congruent gaze condition, participants were found to spend longer looking at the instructor when the gaze cues were more exclusively informative (when provided alongside featural determiners). However, evidence showing that participants actively avoided the instructor when unhelpful gaze cues were used, relative to when no gaze cues were used, could suggest that participants were avoiding a potentially unhelpful reflexive stimulus.

The gaze seeking analysis also provided the only significant effects of AQ group. The AQ group effects must be interpreted with caution, as group assignment was unbalanced within the gaze condition groups. Nonetheless, the analysis of the influence of AQ group and reference type on the proportion of trial time spent fixating the instructor provided intriguing results. Firstly, for the congruent condition only, those in the high AQ group spent more time fixating the instructor on average than those in the low AQ group. This finding could be accounted for by a speculative explanation discussed by Freeth et al. (2013). These experimenters found that low AQ scorers would look more at an individual in a video than high AQ scorers, but that this difference was not present in a real interaction. One possible explanation that Freeth et al. provided was that those with a low AQ score were more sensitive to social cues (so looked more in the video), but were also more sensitive to social norms that prevented them from spending too long looking at the experimenter, in the same way that Laidlaw et al (2011) and Gallup, Hale et al (2012) showed people avoiding looking at others who could potentially look back. If the interpretation of Freeth et al. (2013) is applied to the present study, it could be argued that the low AQ group was more sensitive to the same social norms and therefore spent less time looking at the instructor. A second intriguing finding is that, although featural trials have a greater looking time than spatial trials for both AQ groups, this difference is only significant for the low AQ group. This could be explained by the high AQ group using the congruent cues fairly indiscriminately, whereas the low AQ group were showing sensitivity to social norms and only spending extra time looking at the instructor when the gaze cue was more useful (when provided alongside featural language). These results, however, are based on a small and unbalanced sample of participants, therefore any interpretation must be regarded as speculative.

The selective behaviour toward gaze cues discussed above was also found in the gaze following measures. Immediate gaze following was investigated by measuring the mean proportion of trials in which the first saccade launched after the onset of the gaze cue was in the target direction. When participants in the no gaze condition were provided with featural verbal references they performed no better than chance in this measure. However, they performed significantly better when spatial determiners were used. This can be accounted for by some of these initial eye movements occurring after the end of the sentence: in the spatial determiner sentences, each instruction ended with either the word “left” or “right”, indicating the location of the target. Thus, any participants who made their first eye movement after hearing these instructions were expected to move their eyes in the target direction. There were more first eye movements in the correct direction in the congruent gaze condition relative to the no gaze condition. Pairwise comparisons showed this difference to be significant in both the featural and spatial reference trials, showing evidence that that congruent gaze cues were initially followed by participants regardless of the type of verbal reference used.

The results in the incongruent condition did not differ significantly from those in the no gaze condition. This therefore suggests that the participants did not initially follow anticipated incongruent gaze cues, as the mean proportion of initial eye movements in the correct direction was the same, regardless of whether an incongruent gaze cue or no gaze cue was given. To allow for the possibility that the data in the incongruent condition were skewed by those trials in which the participants did not see the gaze cue, this analysis was repeated using only the trials in which the instructor was fixated at the onset of the gaze cue. On reanalysis, the results did not differ significantly between the no gaze and incongruent conditions. For the incongruent condition only, the next stage of analysis focussed on whether looking at the instructor at the onset of the gaze cue affected the mean proportion of trials with initial eye movements in the correct direction. Participants behaved the same way, regardless of whether or not they looked at the instructor as he provided the gaze cue. This is a clear indicator of selective gaze following behaviour. The incongruent gaze cues had no influence on the direction of the first eye movement after the gaze cue was given, showing that these incongruent (and uninformative) gaze cues were ignored. These results support previous findings showing that unhelpful gaze cues are ignored more (Itier, et al., 2007) and followed less (Macdonald & Tatler, 2013) than those that are helpful.

The time difference between the onset of the sentence and the first fixation on the target object provides an indicator of task performance. In the no gaze condition, participants were slower to fixate the target object when featural determiners were used relative to spatial determiners. This can be explained by the relative ease of the task with spatial references. When the word “left” or “right” was used, half of the objects ceased to be competitors and the search array was halved. No such effect occurred when featural references (“blue”, “red”, etc.) were used. Participants were quicker overall in the congruent gaze condition, where there was no significant difference between performance in the featural and spatial determiner conditions. Pairwise comparisons showed there to be a significant difference between the no gaze and congruent gaze conditions when featural references were used, but not when spatial references were used. This may suggest that gaze cues were utilised more often when they were more informative (when congruent gaze cues accompanied featural references in spoken instructions). However, failure to find a significant difference between the spatial reference trials for the congruent and no gaze conditions may be due to the ease with which the task was completed in the spatial reference trials across both gaze conditions.

While the immediate gaze following and time difference measures show similar patterns of behaviour in the congruent gaze condition relative to the no gaze condition, the relationship differs between the no gaze and incongruent conditions. There was no difference in initial gaze following between these two conditions, however, there was an approaching significant difference between the first fixation time in the no gaze condition between the featural and spatial reference conditions, with the featural trials taking longer. This difference was not found in the incongruent gaze condition. As with the immediate gaze following measure, this dependent variable was reanalysed using only the trials in which the instructor was fixated at the onset of the gaze cue. The difference between reference conditions was found to be significant in the no gaze condition for this reanalysis. However, as with the initial analysis, there was no significant difference between reference conditions in the incongruent gaze conditions. This pattern of difference between the no gaze and incongruent gaze conditions is surprising, given the evidence from the immediate gaze following measure that incongruent gaze cues were ignored. However, these data show that the spatial language advantage seen in the no gaze condition was absent in the incongruent gaze condition. It is unclear why this should be if the cues were actively ignored in the incongruent condition, however, to speculate, this effect may be due to the two pieces of contradictory spatial information provided to the participants when incongruent gaze cues accompanied spatial language: in this condition, the verbal spatial reference ("left" or "right") was accompanied by a gaze cue to the opposite direction. This contradictory gaze cue may have been distracting for the participant and inhibited the beneficial effect of spatial language over featural language, despite participants understanding that the cue was not useful. If this is the reason that the spatial language advantage found in the no gaze condition disappeared in the incongruent gaze condition, then this result may indicate that gaze cues have a non-volitional effect on attention in a real-world interaction, even when gaze cues are used selectively. These findings support the conclusions of lab-based studies that found evidence of automatic and selective gaze seeking (Laidlaw et al, 2012) and gaze following (Ricciardelli et al, 2013) mechanisms. The present study, however, might indicate that these findings can be extended to a real-world environment, in which real gaze cues were used in a real social interaction.

The present study investigated the interaction between gaze cue utilisation and the type of verbal reference used in spoken instructions. The results show that participants selectively sought out gaze cues based on the expected informativeness of these cues, but also that the extent to which these helpful cues were sought out may have been influenced by the social skills of the participant. Evidence for selective gaze following was also found. However, although participants were able to inhibit gaze following when gaze cues were unreliable, there was evidence that the speed of visual search was nevertheless disrupted when inherently spatial gaze cues were accompanied by contradictory spatial verbal references. Previous lab-based studies have suggested both volitional and non-volitional gaze utilisation and the present results now provide evidence for this in a real-world social interaction.

The results showing that gaze cues are utilised based on their perceived informativeness or reliability were found using stimuli with coarse distinctions; gaze cues were either always right, always wrong or not present. Given the difficulties present in real world experiments (lack of experimental control, small sample sizes and time-consuming analysis), the use of coarse distinctions between gaze cueing conditions was necessary to allow for meaningful conclusions. Now that this study has established that cues are selectively used in a real-world search task, based on their relative reliability, the effects of varying gaze cue and language reliabilities can be explored in a more controlled lab-based experiment, using more subtle distinctions in reliability. This will allow for a more detailed understanding of the relationship between language and gaze cue reliabilities.

Chapter 4:

Varying the reliability of gaze and language (i) – Study 3

Introduction

The results from Study 2 show that altering the relative reliability of a gaze cue by changing the type of verbal reference that accompanies it can affect the extent to which the cue is sought out and followed. There is previous evidence that the effects of gaze cues on attention and eye movements can be modulated by the perceived reliability of the cue (Hill et al., 2010). The present study aims to bring together the findings from Study 2 and the methodologies of lab-based studies that have investigated the effects of manipulating cue reliability. A novel paradigm was devised in which both gaze cues and language cues were systematically varied, in order to gain a more detailed understanding of how changing gaze cue and language cue reliabilities affects attention and behaviour.

A reflexive attentional shift (Friesen & Kingstone, 1998) cannot fully account for the way we respond to gaze cue stimuli, as these responses have been shown to vary with perceived cue-reliability. When gaze cue validity is reduced to 20%, the reflexive effect is still present, but only at short SOAs (~100ms) (Kuhn & Kingstone, 2009), suggesting that at later SOAs, a top-down cognitive mechanism overrides the reflexive response. Hill et al. (2010) directly compared responses to different gaze cue reliabilities. Half of their participants took part in a version of the gaze cueing paradigm in which all gaze cues were either valid or invalid. The remaining participants took part in a task in which 50% of gaze cues were valid and the rest invalid. The invalid trials had a detrimental effect on response time for participants in the former task at SOAs up to 150ms only, however in the latter paradigm the detrimental effect was apparent up to 750ms. The authors argued that this was evidence for two streams of attentional control when viewing a gaze cue. Initially there is an automatic orienting effect (present in both tasks) and then a slower, top-down, selective effect (present when cues were 100% invalid). Whether both, one or neither of these attentional effects is unique to social cues is unclear, however it is clear that the perceived reliability of a gaze cue modulates our response.

There is evidence from previous chapters in this thesis, as well as previous literature, that language accompanying gaze cues can affect how these cues are used, even when cues are completely reliable. Knoeferle and Kreysa (2012) found that when unambiguous sentences were harder to process (sentences with an uncommon, but grammatically legal structure), participants were less likely to use helpful gaze cues. Knoeferle and Kreysa argued that the extra cognitive resources required for sentence comprehension leave fewer resources free to utilise non-verbal cues, suggesting a hierarchical use of language and gaze, in which linguistic processing takes precedence over gaze following.

In Study 2 of this thesis, participants were found to look longer at an instructor using helpful gaze cues when the verbal instructions did not provide overlapping spatial information, suggesting that gaze cues are used more when they offer information not provided by language. These language studies suggest that gaze cues are subordinate to spoken language, however, the paradigms described above used gaze in a supportive capacity. The present study employs a task in which the usefulness of gaze cues and language cues are independent of each other, with neither cue subordinate to the other. Each cue directed the participant to one of two potential targets, making language and gaze equivalent cues, except in terms of reliability, which was modulated between sessions.

A highly simplified lab-based version of Study 2 was used to investigate the relationship between gaze and language reliabilities. Participants each carried out 4,320 trials (over nine sessions) of a simple target selection task. The reliabilities of the gaze and language cues were varied, as well as whether targets were referred to with featural or spatial references. The influence of these cues on eye movements and performance was analysed. It is arguable that the effects of language type on gaze following found in Study 2 may simply be a result of the extent to which the gaze cues were sought out. In this paradigm, as in the typical gaze cueing paradigm (Friesen & Kingstone, 1998), participants did not need to seek gaze cues, as the cues appeared on the screen in front of them at the onset of the trial and remained present throughout. Therefore any effect of reference type on gaze following cannot be due to a change in gaze seeking behaviour.

Given the well-established gaze cueing effects using similar paradigms (Friesen & Kingstone, 1998; Ricciardelli et al, 2002), the hypothesis was that gaze cues would influence eye movements even when less reliable than language. However it was also hypothesised that this effect would be modulated by gaze cue reliability, as top-down social-attentional processes inhibit gaze following at low reliabilities (Hill et al, 2010). Previous evidence suggests that gaze cues are ignored in favour of language when gaze provides no additional information (Macdonald & Tatler, 2013) or when available cognitive resources are focussed on language processing (Knoeferle & Kreysa, 2012). Therefore, language cues were predicted to be favoured over gaze cues and lead to task disruption when incongruent.

Method

Participants

Five students (female = 3) aged between 24-28 (mean = 25.2) from the University of Dundee took part in nine one-hour sessions over a two-week period. They each received £20 for their participation.

Materials

Gaze cue stimuli were composed of eight four-second videos. Multiple versions of each cueing video were produced to prevent participants from predicting gaze direction by learning to detect small visual features in the initial frames. The videos began with a face staring forward, then after two seconds the face turned to look left or right. Each of these clips was mirrored to provide 16 unique clips (eight cueing to the left and eight to the right). Language cues were made up of 30 two-second descriptor audio clips ("select the [object name]"), two two-second spatial-determiner audio clips ("on the [left or right]") and twelve two-second featural-determiner audio clips ("that's [colour]"). The end of the noun in the descriptor audio clip always occurred 1,700 ms into the clip and was followed by 300 ms of silence. Each of these descriptor audio clips began with silence, until the instruction sentence began. The duration of the initial silent section varied depending on the duration of the instruction sentence, which varied depending on the duration of the critical noun. The duration of this silence ranged from 28 ms to 802 ms, with a mean duration of 535ms. The determiner audio clips began immediately with the onset of the determiner phrases and the remainder of the two-second clips were filled with silence. As with the initial descriptor clip, the exact duration of these silent sections varied depending on the duration of the particular descriptor phrase. The onset of the determiner word within the determiner phrase occurred on average 172 ms into these clips. On either side of the gaze stimulus there was a target picture. These pictures were 200 × 200 pixels in size and featured objects that differed only in colour (see *Appendix 7* for full list of images). There were 30 pairs of pictures, each of which was used once in each 30-trial block. Stimuli were presented on a 19-inch computer monitor (with a resolution of 1024 × 768), approximately 64 cm away from the participant, whose head was stabilised on chin and forehead rests. At this distance the screen was 31.8° × 23.8° of visual angle. A control pad with a left and a right trigger was used for participant responses.

Eye movement recording

This experiment was carried out using an SR Research EyeLink 1000 desk-mounted eye-tracker and the SR Research Experiment Builder software. This system uses corneal reflection and pupil position to calculate where a participant is fixating. Calibration involved the participant fixating on nine markers on the screen. Once calibrated, a verification procedure took place. If the verification procedure found mean spatial accuracy error to be more than .5°, or if any one of the spatial accuracy errors was greater than 1°, calibration and verification procedures were repeated. Verification (and if necessary, re-calibration) was carried out after every block of 30 trials. The mean calibration error was .369° (SD = .112°) of visual angle. Before each trial, participants fixated a marker in the middle of the screen. The lead experimenter could see the estimated fixation point on their display and was required to accept this fixation in order for the trial to begin. The average error for the single-point calibration check before each trial was .502° (SD = .260°) of visual angle.

Design

The present experiment used a within-subjects design. Between sessions there were two independent variables: The probability that the gaze cue was correct (50%, 80%, 100%) and the probability that the language cue was correct (50%, 80%, 100%). Each of the nine sessions used a unique combination of these probabilities. Within each session, there were 16 blocks of trials and in each of these the proportion of correct cues matched the proportion for the whole session. Within each block whether each cue was correct varied (except when the cue was 100%). Half of the language cues in each block used spatial references (left or right) and the other half used featural references (colour).

Procedure

Each participant was initially informed that they would be required to take part in nine sessions over a two-week period. The lead experimenter explained to the participant that in each session they would have to do the same thing; complete 480 trials of a simple decision task. Participants were informed that in each trial they would be shown pictures of two objects on a screen, one on the left and one on the right, and that they would have to choose between these by looking at their chosen target and then pressing the left or right trigger on the control pad. They were told that there would be verbal and non-verbal cues to help guide them, but that these might not always be reliable. Participants were then informed that they would receive immediate visual feedback after each response and that they should try their best to get as many correct as possible. The eye-tracker was then set-up, calibration carried out, and then the first trial began. The trial started with one of the 16 four-second videos playing in the centre of the screen. Throughout the trial the target items were displayed at either side of the video. Alongside the video, one of the two-second descriptor audio clips played. When the audio clip finished (simultaneously with the onset of the gaze cue), the audio determiner clip began playing (Figure 4.1). After this point, participants were able to select a target, after which they were informed (on screen) whether they were "correct" (in green) or "wrong" (in red). Participants were free to take breaks at any time. The session was complete after 480 trials. Participants returned on eight more occasions to repeat this procedure. Across the sessions language cue reliability (50%, 80%, 100%) and gaze cue reliability (50%, 80%, 100%) varied in all possible combinations. Each of the 16 video stimuli was used 30 times, once with each object. Participants were not told the reliabilities of the gaze and language cues.

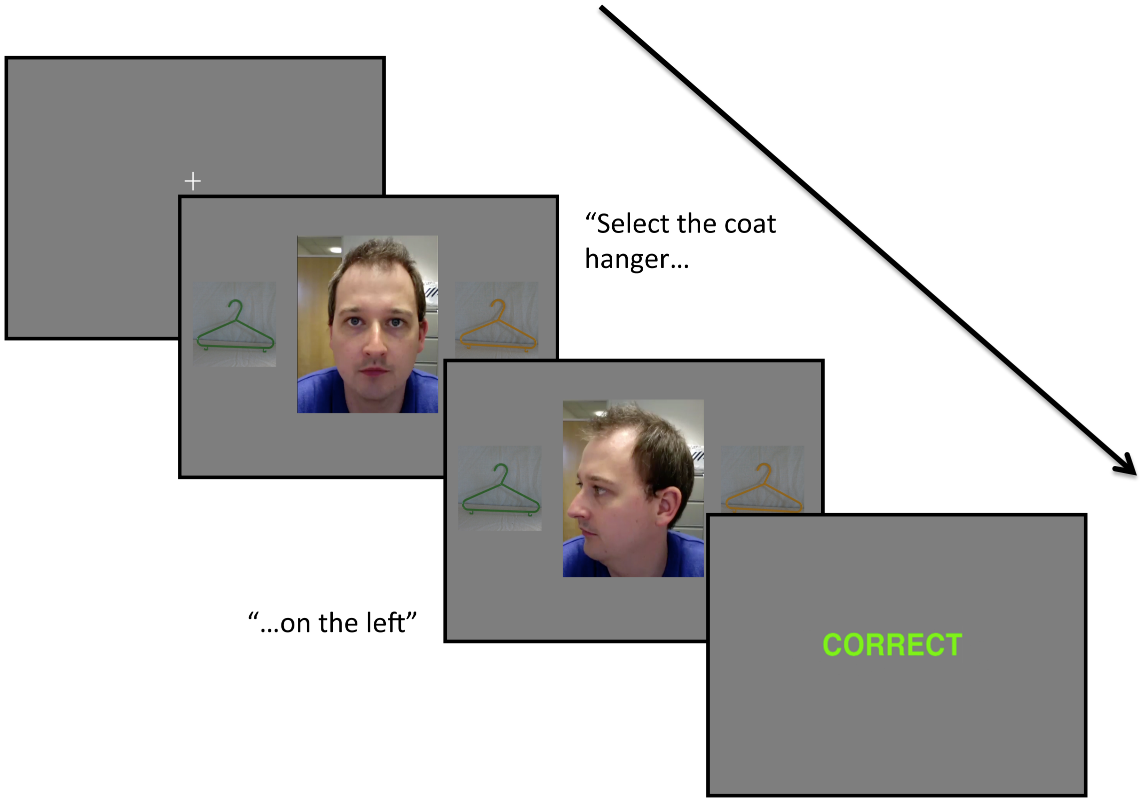

Figure 4.1. Outline of a trial. After the central fixation point was fixated, a video began showing a face looking at the participant, with pictures on either side that differed in colour. While this video played the first part of the sentence was heard. After two seconds the head began to turn towards one of the targets while the second part of the sentence was heard. After making a decision the participant received immediate feedback

Analysis

Five main dependent variables were used for the analysis of the results: two performance measures (accuracy and response time) and three eye movement measures (first saccade direction, first saccade latency and time to first fixate the target). The initial analysis focused on how these variables were affected by the overall reliability of both language and gaze cues as well as the type of reference used. For analysis, 3 (gaze reliability) × 3 (language reliability) × 2 (reference type) ANOVAs were performed for each of the dependent variables.

The interaction between cue congruity and reliability was also investigated. To do this the session in which both cues were 50% reliable and the session in which both cues were 80% reliable were analysed. These sessions were used because they were the only sessions with incongruent trials in which both cues were equally reliable. Secondly, sessions in which one cue was 100% reliable and the other cue less reliable were analysed to see if there was a detrimental effect of incongruent gaze and language cues and to investigate whether this was mediated by cue reliability. For many of these analyses uneven sample sizes were compared, so traditional ANOVA models were avoided. Instead, the *lme4* package in the R statistical programming environment was used to run linear mixed effects (LMM) models. The approach to LMMs in this chapter was identical to my approach in Chapter 3. As with earlier chapters, pairwise comparisons were analysed with Tukey-tests using the *glht()* function in the *multcomp* library (Hothorn, Bretz & Westfall, 2008). Some additional simple comparisons were carried out using independent samples *t*-tests.

Results

Effect of changing cue reliability

Each of the nine sessions used a different combination of language and gaze cue reliabilities. Figure 4.2 shows the results for accuracy in each session. A clear interaction can be seen between language and gaze cue reliability. A 3×3×2 ANOVA confirmed main effects of language reliability, $F(2,72) = 219.68$, $p < .001$, and gaze reliability, $F(2,72) = 214.76$, $p < .001$, as well as a significant interaction, $F(4,72) = 72.40$, $p < .001$, but no three-way interaction with reference type, $F(4,72) = .274$, $p = .893$. There was also no main effect of reference type, $F(4,72) = .095$, $p = .759$, nor any interactions with either gaze reliability, $F(4,72) = .468$, $p = .628$, or language reliability, $F(4,72) = .158$, $p = .854$. Accuracy approached ceiling whenever at least one cue was 100%, approached .8 when the most reliable cue was 80% accurate and approached .5 when both cues were 50% accurate. One participant who behaved atypically can account for the large error bar in the session in which gaze cues were 80% reliable and language cues were 50% reliable.

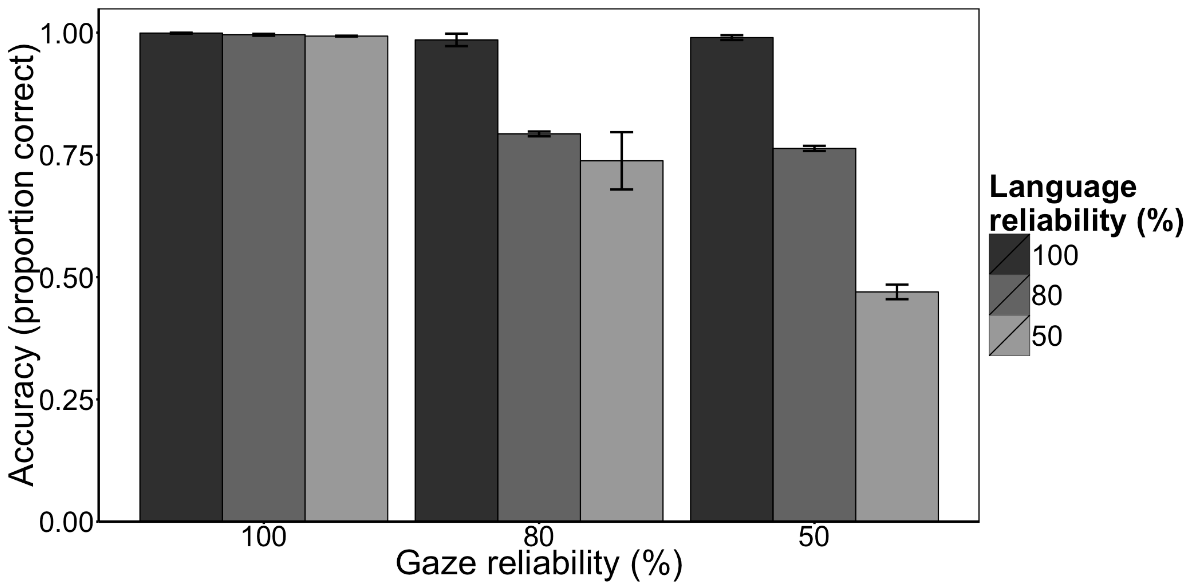

Figure 4.2. The mean accuracy (proportion of correct responses) across all gaze and language reliabilities. Error bars show standard error of means.

The proportion of first saccades in the correct direction (Figure 4.3) was found to decrease as gaze reliability decreased, $F(2,72) = 630.13$, $p < .001$. There was no significant effect of language reliability, $F(2,72) = .134$, $p =.874$, however there was a significant interaction between gaze and language reliabilities, $F(4,72) = 3.206$, $p = .018$. When gaze was 80% reliable, the proportion of correct first saccades increased with decreasing language reliability. However, when gaze was 50% reliable, the proportion of correct first saccades increased with increasing language reliability. There was no effect of reference type, $F(2,72) = .010$, $p = .921$, nor was there any interactions found with gaze reliability, $F(4,72) = .008$, $p = .992$, or language reliability, $F(4,72) = .358$, $p = .700$, nor between all three factors, $F(4,72) = .165$, $p =.955$.

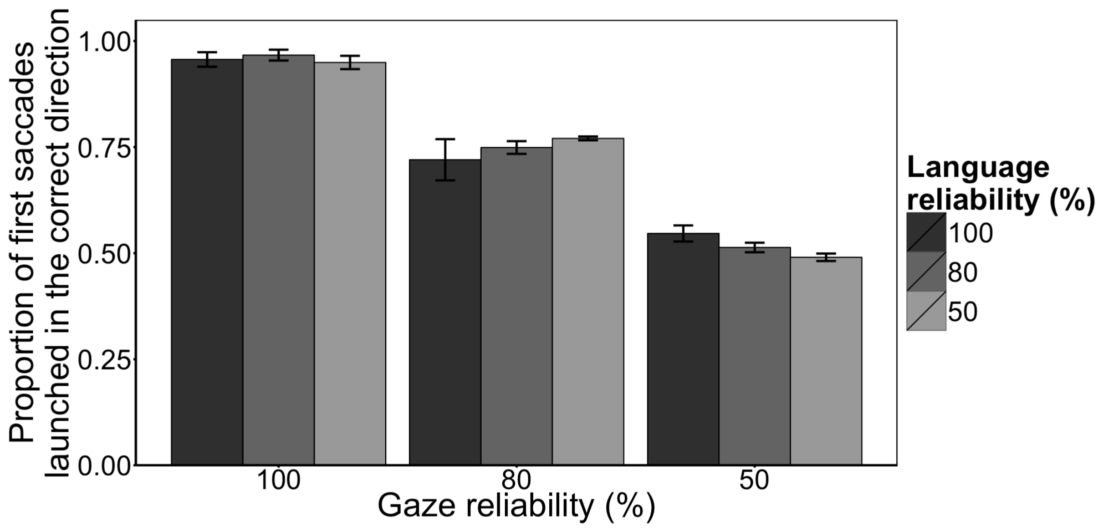

Figure 4.3. The mean proportion of first saccades in the correct direction across all gaze and language reliabilities. Error bars show standard error of means.

The mean first saccade latency is shown in Figure 4.4. As gaze reliability decreased, time to launch the first saccade increased, $F(2,72) = 12.062$, $p < .001$. There was no main effects of language reliability, $F(2,72) = 1.536$, $p = .222$, or reference type, $F(2,72) = .133$, $p = .716$. Gaze reliability did not significantly interact with language reliability, $F(4,72) = 1.441$, $p = .229$, or reference type, $F(4,72) = .072$, $p = .931$, nor was there a three-way interaction, $F(2,72) = .007$, $p > .999$. There was also no interaction found between language reliability and reference type, $F(2,72) = .040$, $p = .961$.

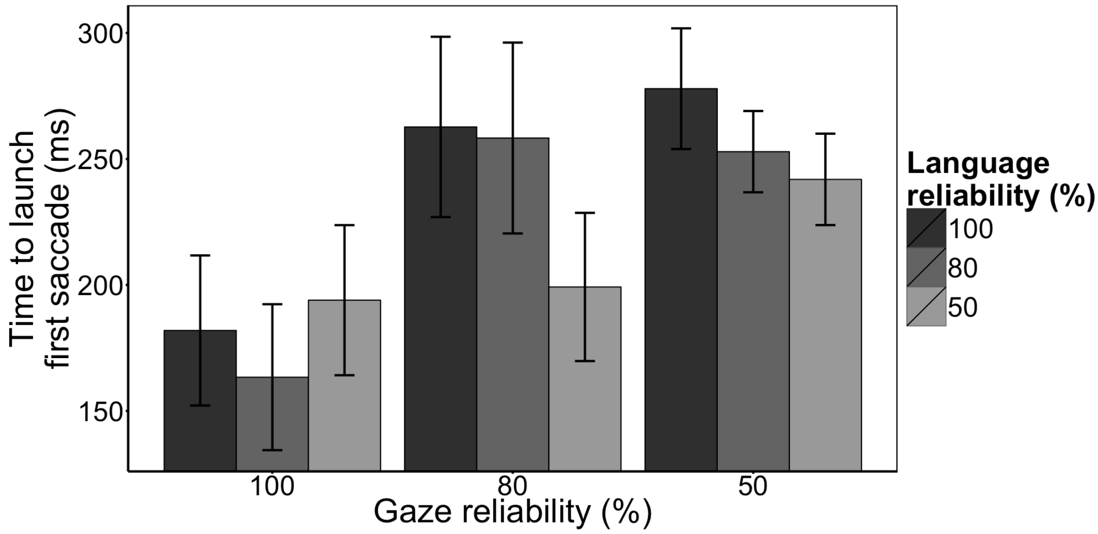

Figure 4.4. The mean first saccade latency across all gaze and language reliabilities. Error bars show standard error of means.

The time to first fixate the target (FT) (Figure 4.5) increased as gaze reliability decreased, $F(2,72) = 34.43$, $p < .001$. In Figure 4.5, FT appears quickest for low reliability language conditions, except when gaze was 100% accurate. A significant interaction between gaze and language reliability, $F(4,72) = 3.66$, $p = .009$, as well as a main effect of language reliability was found, $F(2,72) = 4.25$, $p = .018$. As with the accuracy measure, there was no three-way interaction, $F(4,72) =.019$, $p = .999$, nor was there a main effect of reference type, $F(2,72) = .420$, $p = .519$. No significant interaction was found between reference type and gaze reliability, $F(4,72) = .025$, $p = .976$, or language reliability, $F(4,72) = .023$, $p = .978$.

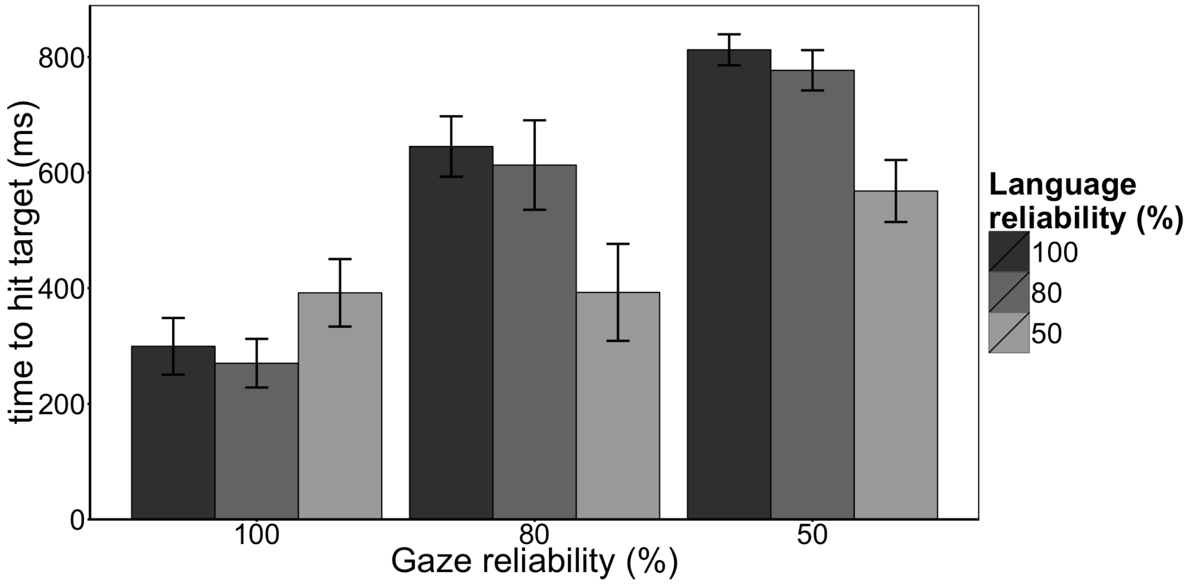

Figure 4.5. The mean time to first fixate the target (FT) across all gaze and language reliabilities. Error bars show standard error of means.

Response time similarly increased as gaze reliability decreased, $F(2,72) = 13.89$, $p < .001$ (Figure 4.6). There was no main effect of language reliability, $F(2,72) = .073$, $p = .930$, nor any interaction between gaze and language reliability, $F(4,72) = 1.19$, $p = .321$. There was no significant three-way interaction or main effect of reference type, $F(4,72) = .019$, $p = .999$. Additionally reference type did not interact with either gaze, $F(4,72) = .027$, $p = .973$, or language reliability, $F(4,72) = .016$, $p = .984$.

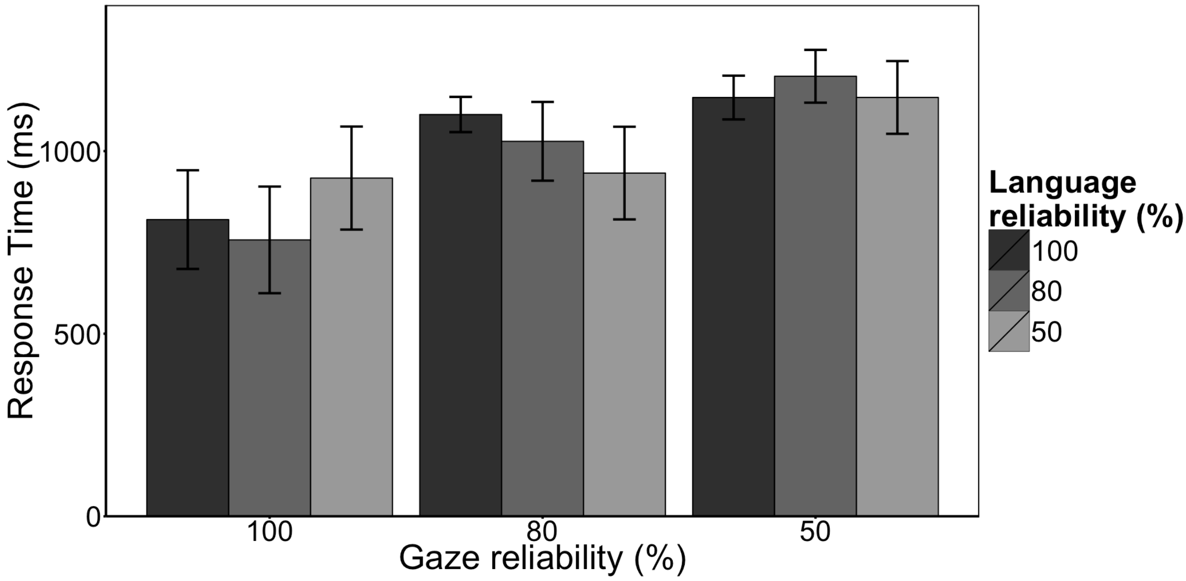

Figure 4.6. The mean response time (RT) across all gaze and language reliabilities. Error bars show standard error of means.

Effects of cue congruity

Effects of equally reliable incongruent cues. Gaze cues and language cues are both naturally informative cues that have been manipulated to be equally useful in certain sessions in this paradigm. A bias for selecting the target cued by either stimuli when they were equally reliable may be indicative of a natural preference for one cue over the other.

Out of the three sessions in which language and gaze cues were equally reliable, two included trials with gaze and language cues incongruent to each other. The first of these was the session with 50% reliable gaze and language cues. In this session, neither cue was at all informative, therefore participants performed no better than chance (Figure 4.2). Figure 4.7 shows the proportion of incongruent trials in which participants selected the picture cued by language and gaze. When cues were incongruent with eachother, there was no significant difference found between the proportion of targets selected that were cued by gaze and language, $t(8) = .098$, $p = .924$.

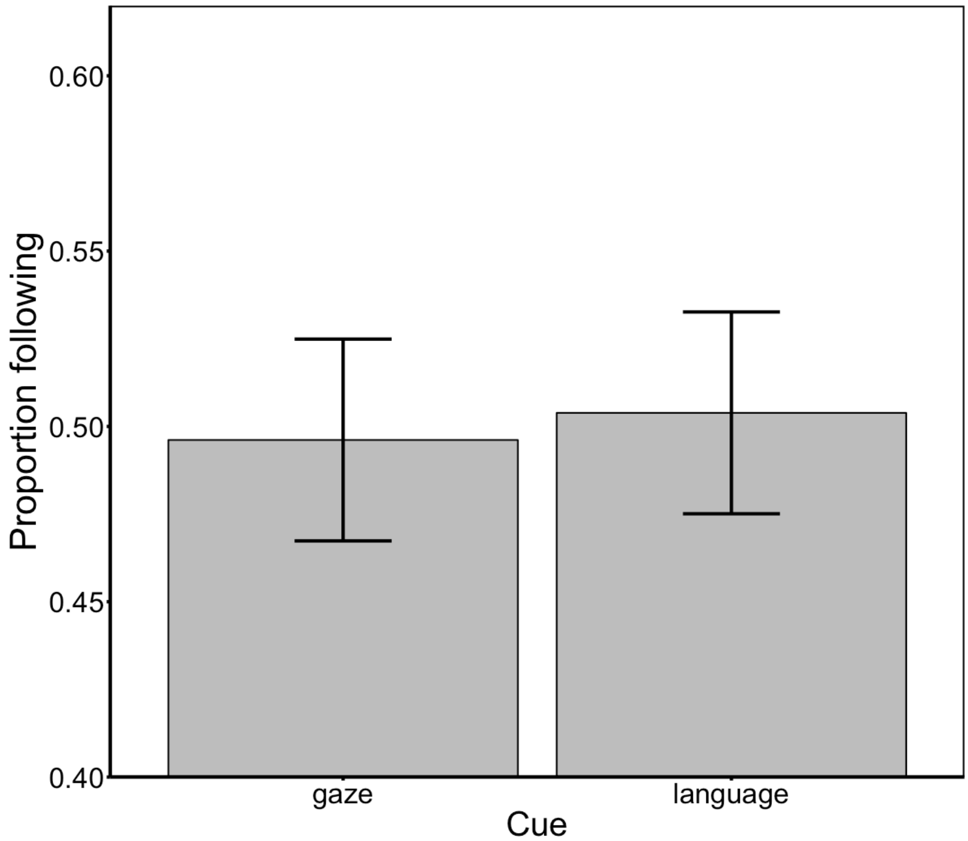

Figure 4.7. The proportion of targets selected that were cued by gaze and language in 50% gaze and language cue reliability conditions when cues were incongruent. Standard error across trials is shown.

A lack of a significant difference between the proportion of targets cued by gaze and targets cued by language that were selected may indicate that participants were not naturally biased to either cue. However, all five participants reported "giving up" during this session. All participants completed all trials, but since neither cue was any more informative than chance, there was no beneficial strategy for participants to employ, leading to participants losing motivation to get as many trials correct as possible.

Participants were much more likely to remain motivated during the session with 80% gaze and language reliability. This is because both cues were mostly correct, so participants could employ one of two equally advantageous strategies: follow gaze cues or follow language cues. One participant was removed from this analysis as they exhibited very different behaviour to the other participants. Figure 4.8 shows the proportion of targets cued by gaze and language that were selected, for trials in which language and gaze indicated different objects (incongruent trials). Language cues were selected in significantly more incongruent trials, $t(6) = 8.627$, $p < 0.001$. This indicates that when cues were equally reliable and incongruent with each other, participants were more likely to select the target cued by language.

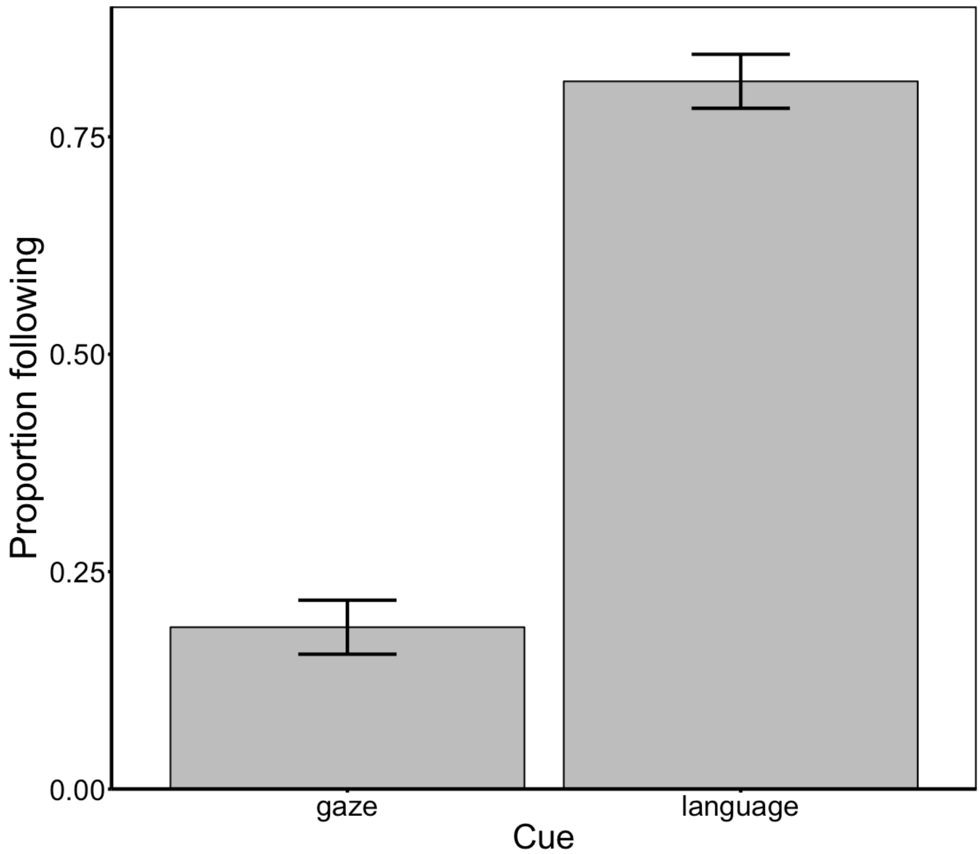

Figure 4.8. The proportion of targets selected that were cued by gaze and language in the 80% gaze and language cue reliability conditions when cues were incongruent. Standard error across trials is shown.

To investigate if the type of language had any influence on the preference for selecting language-cued targets over gaze-cued targets, two LMMs were carried out on the proportion of incongruent trials in which participants selected the language cue. The first of these models (Model 4.1) (using reference type as a fixed factor and subject and item as random factors) was fit to data from the session with 50% reliable language and gaze cues (Figure 4.9a). No effect of reference type was found on the proportion of trials in which the participant selected the language cue, β = -.221, *SE* =.184, *t* = -1.200, *p* = .230.

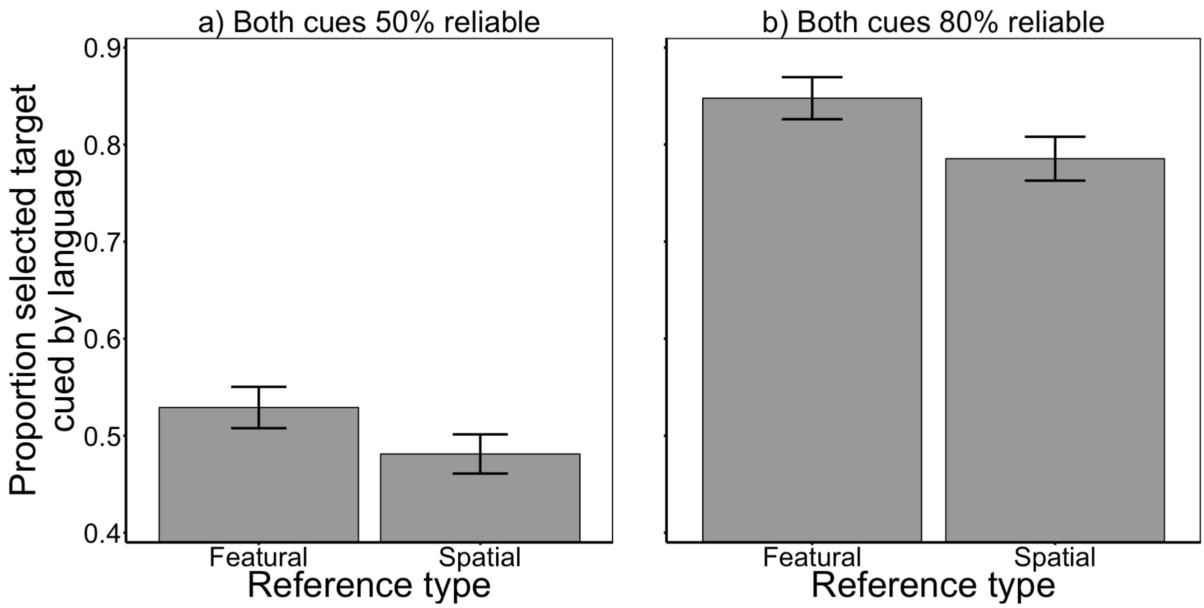

Figure 4.9. The proportion of targets selected that were cued by gaze and language in featural and spatial reference trials. Results are shown for incongruent trials in the sessions with a) 50% gaze and language reliability and b) 80% gaze and language reliability.

The second LMM (Model 4.2) also used reference type as a fixed factor (and subject and item as random factors), and found that in the 80% gaze and language reliability session, participants were more likely to select the target cued by language over the target cued by gaze if a featural reference was used, rather than a spatial reference (Figure 4.9b). This difference was found to be approaching significance, β = -.551, *SE* =.301, *t* = -1.829, *p* = .067.

Effects of the less reliable cue. An analysis of the four sessions in which one cue was 100% accurate and the other cue less accurate was used to investigated the distracting effect of less reliable cues. Specific focus was on whether the reliability of the less-reliable cue affected the measures, even though the other cue was completely reliable. The optimal strategy in these sessions involved ignoring the less reliable cues, so any effect of these subordinate cues is indicative of non-volitional capture of attention. In these four sessions, there were 2-levels of congruity: 1) cues congruent with each other and correct and 2) cues incongruent with each other, with the less reliable cue incorrect.

Figures 4.10a and 4.10b show the accuracy results for both congruent and incongruent trials in these four sessions. An LMM (Model 4.3) of accuracy (using congruity and gaze reliability as fixed factors and subject and item as random factors) for trials where language reliability was 100% (Figure 4.10a) showed that participants were significantly more accurate in congruent trials, $\beta = 2.099$, $SE = .487$, $z = 4.312$, $p < .001$. There was a significant interaction between congruity and gaze reliability, $\beta = 2.856$, $SE = .629$, $z = 4.541$, $p < .001$, although there was no overall effect of gaze reliability, $\beta = .811$, $SE = .669$, $z = 1.213$, $p = .225$. Post-hoc Tukey tests showed that the difference between incongruent and congruent trials was significant when gaze-cues were 80% reliable, $p < .001$, but not when gaze cues were 50% reliable, $p = .294$.

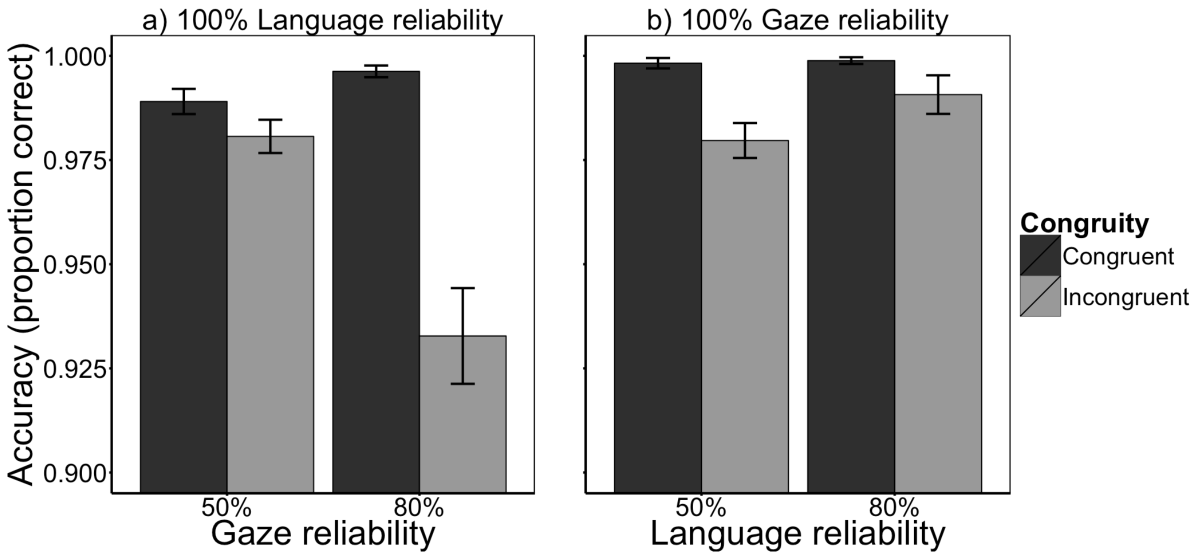

Figure 4.10. The mean accuracy (proportion) for participants in congruent and incongruent cue trials for a) sessions with 100% language reliability and 50% and 80% gaze reliabilities and b) sessions with 100% gaze reliability and 50% and 80% language reliabilities.

An LMM (Model 4.4) (using congruity and gaze reliability as fixed factors and subject and item as random factors) of accuracy for trials in which gaze reliability was 100% (Figure 4.10b) showed that congruent trials have significantly higher accuracy scores than incongruent trials, $\beta = 2.469$, $SE = .735$, $z = 3.361$, $p < .001$. Language reliability did not have an overall effect, $\beta = .778$, $SE = .658$, $z = 1.182$, $p = .237$, nor was there a significant interaction between the two factors, $\beta = .364$, $SE =1.140$, $z = -.319$, $p = .750$.

It is possible that the interaction present in Figure 4.10a may be caused by the difference in the overall number of incorrect gaze cues between the two gaze reliability conditions. It may be that in both sessions, roughly the same number of mistakes was made, but that the lower number of incongruent gaze cue trials in the

80% gaze condition led to a lower proportion of correct responses when gaze cues were incongruent. To explore this possibility, the proportion of correct responses in the first two blocks of the 50% gaze session were compared to the proportion of correct responses in the first five blocks of the 80% gaze session (Figure 4.11a). These sections of each session contained 30 incorrect trials, thus the total number of incorrect trials could not skew the proportion of correct responses. An LMM (Model 4.5) of accuracy for the initial blocks (using congruity and gaze reliability as fixed factors and subject and item as random factors) found an effect of congruity, β = 2.434, *SE* = .937, z = 2.596, p = .009, but no effect of gaze reliability, β = -.334, *SE* = .975, z = -.342, p = .732, nor any interaction, β = 2.294, *SE* = 1.571, z = 1.460, p = .144. Despite the interaction not reaching significance, there was *a priori* interest in whether cue incongruity had different effects in the 50% and 80% gaze reliability conditions. Post-hoc Tukey tests showed that cue incongruity had a significant detrimental effect on accuracy in the 80% gaze reliability condition, $p < .001$, but not the 50% gaze reliability condition, p = .852.

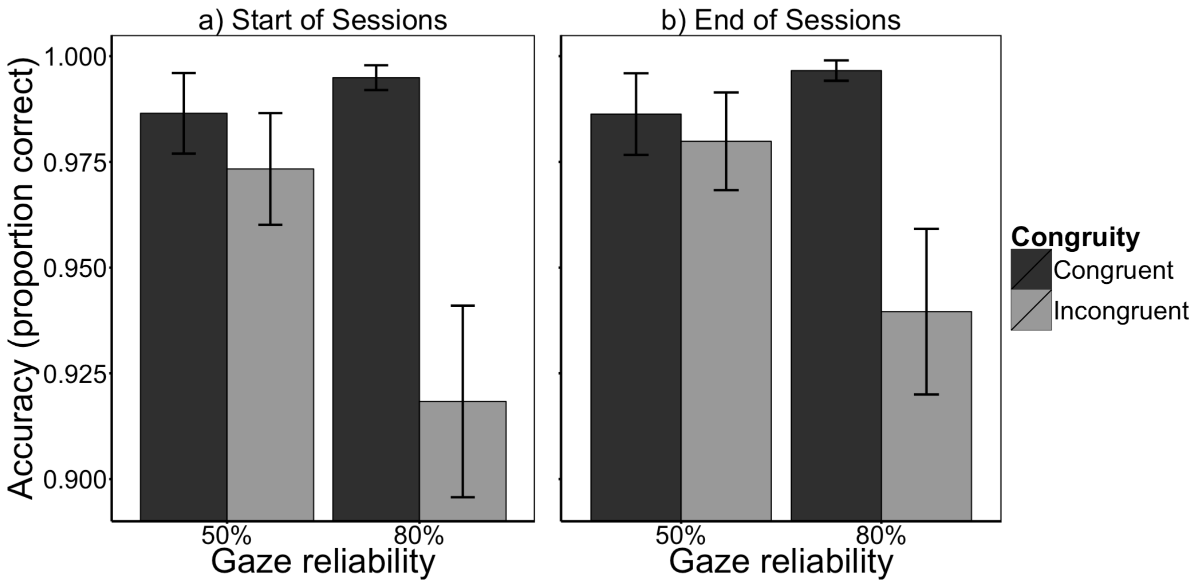

Figure 4.11. The mean accuracy (proportion) for participants in congruent and incongruent cue trials for sessions with 100% language reliability and 50% and 80% gaze reliabilities. Results are shown for a) the first and b) the last five blocks of the 80% gaze condition and two blocks of the 50% gaze condition.

To investigate the possibility that the detrimental effect of incongruent gaze cues in the 80% reliability condition may have been short-lived, the proportion of correct responses in the final two blocks of the 50% gaze session was compared to the proportion of correct responses in the final five blocks of the 80% gaze session (Figure 4.11b). An LMM (Model 4.6) of accuracy for the initial blocks (using congruity and gaze reliability as fixed factors and subject and item as random factors) found an effect of congruity, $\beta = 2.398$, $SE = .685$, $z = 3.502$, $p < .001$, but no effect of gaze reliability, $\beta = .539$, $SE = .779$, $z = .693$, $p = .688$, nor any interaction, $\beta = 2.675$, $SE = 1.744$, $z = 1.534$, $p = .125$. As with the analysis of the initial blocks, post-hoc Tukey tests showed that cue incongruity had a significantly detrimental effect on accuracy in the 80% gaze reliability condition, $p < .001$, but not the 50% gaze reliability condition, $p = .989$.

The proportion of first saccades after the onset of the cues (exactly 2 seconds into the video) that were in the direction of the target is shown in Figures 12a and 12b. For the two sessions with 100% reliable language cues and less reliable gaze cues, an LMM (Model 4.7) (using congruity and gaze reliability as fixed factors and subject and item as random factors) showed no significant effect of gaze reliability, $\beta = -.043$, $SE = .089$, $z = -.488$, $p = .626$, but a significant effect of congruity, $\beta = 3.921$, $SE = 1.194$, $z = 3.285$, $p = 001$, and a significant interaction, $\beta = .665$, $SE = .183$, $z = 3.633$, $p < 001$ (Figure 4.12a). Pairwise comparisons showed that there was no significant difference between the proportion of initial saccades in the correct direction between the 80% and 50% gaze cue reliability sessions when gaze cues were congruent ($p = .360$), but there was a significantly higher proportion of correct first saccades in the 50% gaze reliability condition when gaze cues were incongruent ($p = .015$).

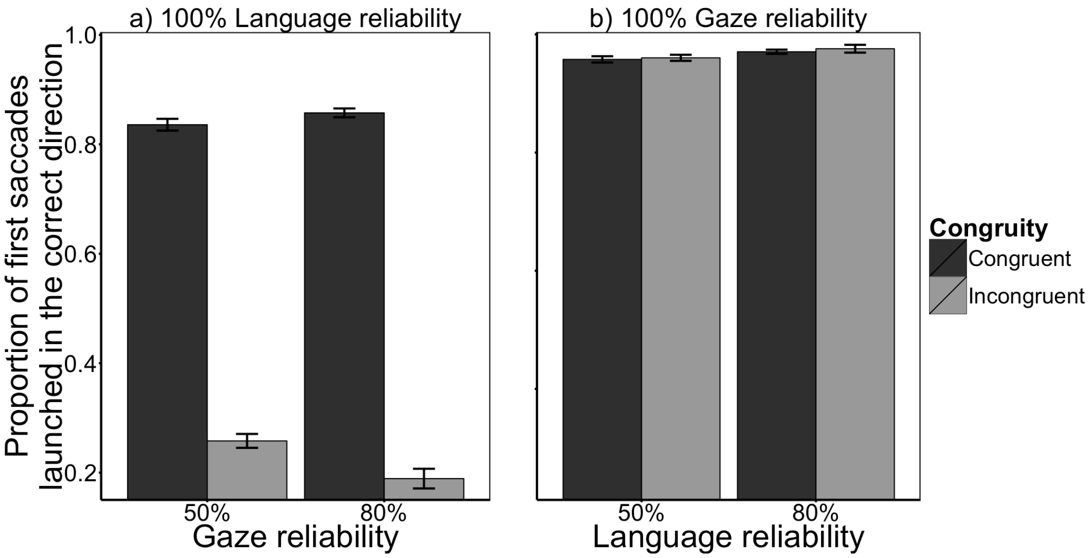

Figure 4.12. The proportion of first saccades launched in the correct direction after the onset of the cues for participants in congruent and incongruent cue trials for a) sessions with 100% language reliability and 50% and 80% gaze reliabilities and b) sessions with 100% gaze reliability and 50% and 80% language reliabilities

An LMM (model 4.8) of proportion of correct first saccades for trials in which gaze reliability was 100% (Figure 12b) showed no overall difference between the proportion of initial saccades launched in the correct direction in congruent and incongruent trials, β = -.075, *SE* = .192, z= -.392, *p* = .695. Likewise, there was no significant difference between sessions in which language reliability was 80% compared to 50%, β = .576, *SE* = .342, *z* = 1.683, *p* = .092, nor was there any interaction between these two factors, β = -.150, *SE* = .362, *z* = -.415, *p* = .678.

LMMs of first saccade latency were carried out for these four sessions (Figures 4.13a and 13b). An LMM (Model 4.9) of saccade latency (using congruity and gaze reliability as fixed factors and subject and item as random factors) for trials in which language reliability was 100% (Figure 4.13a) showed no overall difference between the mean saccade latency in congruent and incongruent trials, β = .660, SE = 7.693, t= .086, p = .923, nor was there a significant difference between sessions in which gaze reliability was 80% compared to 50%, β = -15.627, SE = 24.757, t = -.631, p = .493. No interaction was found between these two factors, β = -23.256, SE = 21.812, t = -1.066, p = .285.

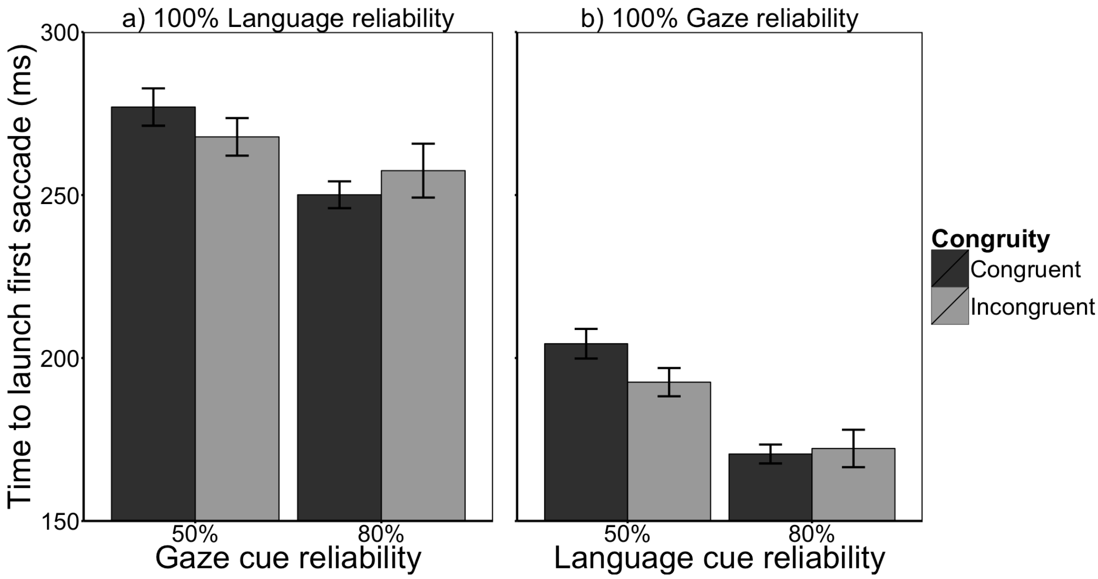

Figure 4.13. The mean first saccade latency (ms) for participants in congruent and incongruent cue trials for a) sessions with 100% language reliability and 50% and 80% gaze reliabilities and b) sessions with 100% gaze reliability and 50% and 80% language

An LMM (Model 4.10) of first saccade latency for trials in which gaze reliability was 100% (Figure 4.13b) showed no overall difference between the mean saccade latency in congruent and incongruent trials, β = 5.343, SE = 4.913, t = -1.088, p = .287. However, mean saccade latency was significantly longer when language reliability was 50% compared to 80%, *β* = -31.976, *SE* = 11.020, *t* = -2.902, *p* = .017. There was no significant interaction between these two factors, *β* = -12.363, *SE* = 10.473, *t* = .-1.181, *p* = .250.

Figures 4.14a and 4.14b show the time to first fixate the target (FT) for congruent and incongruent trials in the four sessions. An LMM (Model 4.11) of FT (using congruity and gaze reliability as fixed factors and subject and item as random factors) for trials where language reliability was 100% (Figure 4.14a) showed participants were significantly quicker to first fixate the target in congruent trials, *β* = -277.04, *SE* = 39.60, *t* = -6.996, *p* < 001. There was an approaching significant interaction between congruity and gaze reliability, *β* = 230.48, *SE* = 112.70, *t* = -2.045, *p* = .059, although there was no overall effect of gaze reliability, *β* = 85.00, *SE* = 67.86, *t* = -1.253, *p* = .198. Post-hoc Tukey tests showed significant differences in the time to first fixate the target between the gaze cue reliability conditions for both congruent and incongruent cues. However, the time to first fixate the target was shorter in the 80% reliable condition, relative to the 50% reliable condition, for congruent cues, *p* < .001, but longer for incongruent cues, *p* < .001.

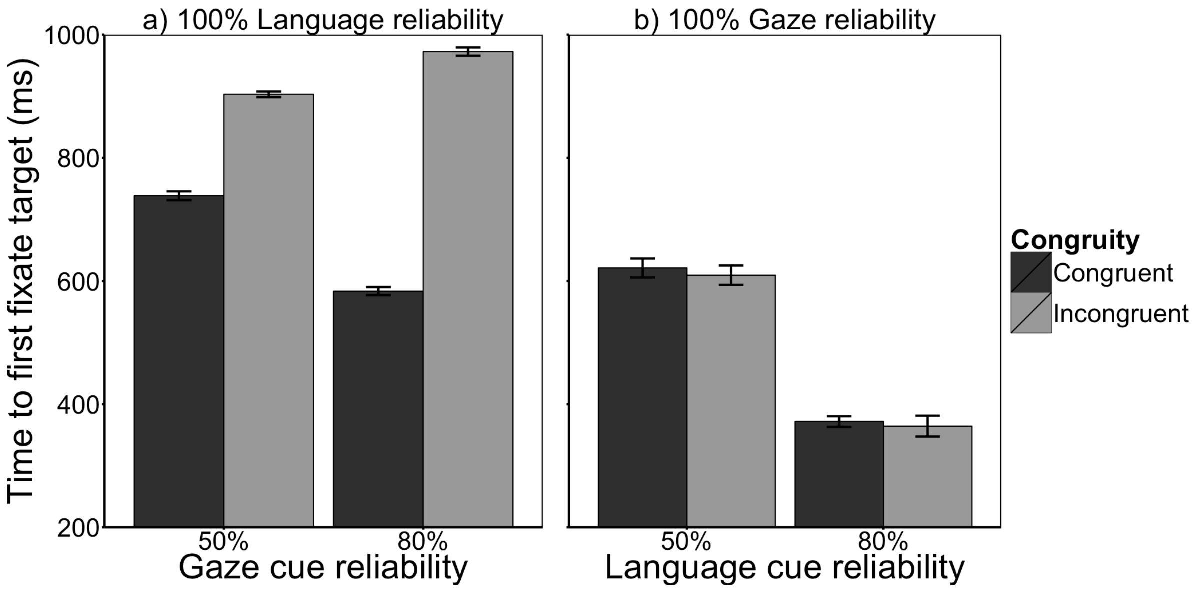

Figure 4.14. The mean time to fixate target (ms) for participants in congruent and incongruent cue trials for a) sessions with 100% language reliability and 50% and 80% gaze reliabilities and b) sessions with 100% gaze reliability and 50% and 80% language reliabilities

An LMM (Model 4.12) of FT for trials in which gaze reliability was 100% (Figure 4.14b) showed no overall difference between the time to first fixate the target in congruent and incongruent trials, $\beta = 4.176$, $SE = 7.031$, $t = .594$, $p = .553$. However, the time to first fixate the target was significantly quicker when language reliability was 80% compared to 50%, $\beta = -123.067$, $SE = 56.241$, $t = -2.188$, $p = .047$. There was no interaction between these two factors, $\beta = .607$, $SE = 12.532$, $t = .048$, $p = .965$.

Finally, LMMs of response time (RT) were performed for trials in which language cues were 100% reliable (Figure 4.15a). Congruent trials were found to have significantly quicker response times (Model 4.13), $\beta = 115.660$, $SE = 21.440$, $t =$ -

5.395, p = .001, and an approaching significant interaction was found between congruity and gaze reliability, β = -100.350, SE = 48.980, t = -2.049, p = .057. There was no overall effect of gaze reliability, β = -10.860, SE = 39.790, t = -.273, p = .761. Post-hoc Tukey tests showed that the incongruent trials had significantly longer response times when gaze cue reliability was 80% compared to 50%, p < .001, and that congruent trials had a significantly longer response time when gaze cue reliability was 50% compared to 80% p <.001.

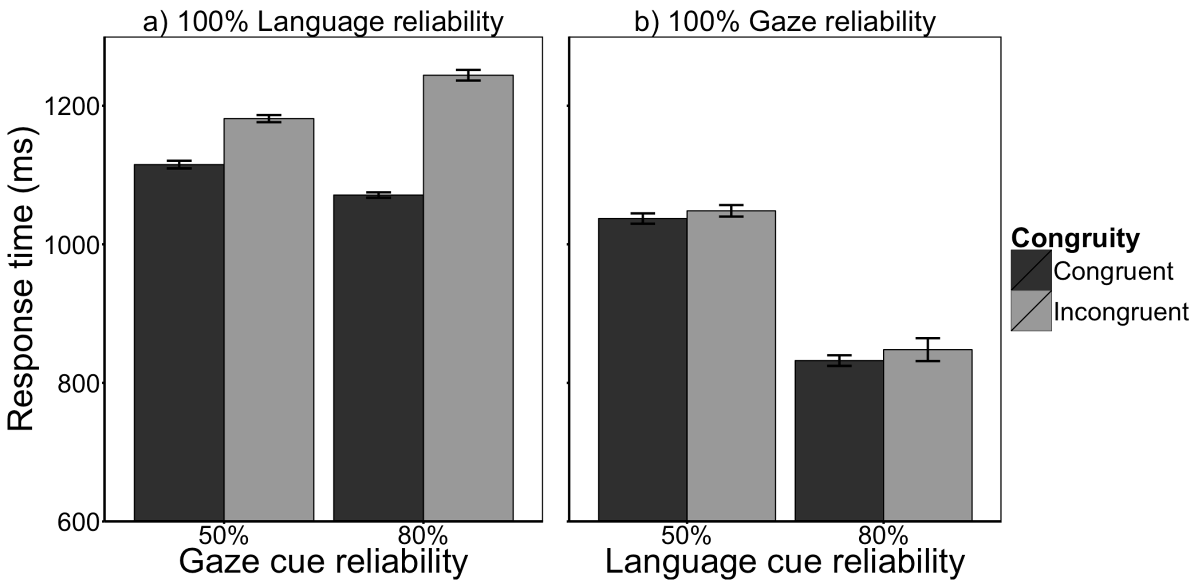

Figure 4.15. The mean response time (ms) for participants in congruent and incongruent cue trials for a) sessions with 100% language reliability and 50% and 80% gaze reliabilities and b) sessions with 100% gaze reliability and 50% and 80% language

When gaze reliability was 100% (Figure 4.15b) there was no difference between response times in congruent and incongruent trials (Model 4.14), β = -7.601, SE =

10.857, t = -.700, p = .567. Mean response was quicker when language reliability was 80% compared to 50%, but this difference was not significant, β = -166.698, SE = 101.074, t = -1.649, p = .108. There was no interaction between these two factors, β = -6.906, SE = 13.633, t = -.507, p = .713.

Discussion

The aim of this study was to investigate how varying the reliability of language cues and gaze cues affects attention and performance in a simple task. The previous real world experiments in this thesis necessarily used relatively coarse-grained measures to investigate this relationship. Using a lab-based desk-mounted eye-tracking paradigm, manipulations and measures were more fine-grained in the present study, allowing for a more detailed investigation of the relationship between gaze and language reliabilities. A target selection task was used in which the accuracy of both gaze and language cues varied across sessions. In order to gather reliable data each session comprised 480 trials, allowing participants to learn cue reliabilities. Additionally, the sessions were conducted on different days, to reduce the chance of strategies learned in one session affecting another. Participants generally learned quickly and used the most informative cue provided to inform their response. Gaze reliability had an effect on the time to first fixate the target (FT) and response time (RT), showing that these cues speed up performance when useful. Gaze and language cues had rather different effects on attention and performance when they were the less reliable cue. Language was most disruptive when least reliable, suggesting that it was the favoured cue (Knoeferle & Kreysa, 2012). Gaze cues were followed more often when more reliable, supporting the findings of Macdonald and Tatler (2013).

The analysis of the effects of cue reliability on accuracy shows a clear interaction between the language and gaze cues. Participants strategically followed the instructions of the most reliable cue at their disposal. The time to hit target (FT) was also affected by both gaze and language reliabilities, and these effects interacted, indicating that the relative usefulness of these cues affected eye movement behaviour. Mean response time was not significantly influenced by language reliability, but was affected by gaze reliability. This effect could be due to the incorrect gaze cues guiding attention to the incorrect item before participants reoriented attention to the correct item. This idea is supported by the first saccade direction measure, which showed that when gaze cues were 50% reliable, participants launched their first saccade in the wrong direction in around 50% or trials, whereas when gaze cues were 80% reliable, their first saccades were in the wrong direction in around 20% of trials, indicating that initially, gaze cues are persistent at low reliabilities. However, the findings from the first saccade latency measure, suggest that the effect of gaze reliability on response time may also have been partly due to the participants actively avoiding following these cues when they were less reliable and instead waiting until they heard the language cue before moving their eyes to the correct target. The first saccade latency measure shows that the lower gaze cue reliability, the longer participants waited to launch their first saccade. The mechanism for this finding is unclear. To speculate, this effect may be due to increased uncertainty of the gaze cues leading to an overall increase in hesitancy in the task. Alternatively, these means may have been skewed by more trials in which participants waited for language to identify the target when gaze cues were less reliable.

In all of the main measures, reference type had no significant effect, nor did it interact with either gaze or language reliability. In this paradigm at least, there was no evidence that the type of language used had any effect on whether gaze or language cues were followed. This finding is in stark contrast to the findings of Study 2, which showed that varying the type of language used in an instruction varied gaze cue utilisation. This may be because the effects found in Study 2 were confined to the gaze-orienting stage of gaze cue utilisation, and therefore when this stage was removed, no effect remained. Alternatively, it is possible that in this simple task, the spatial location of a target was very quickly ascertained from a featural determiner. In the featural reference condition in Study 2, a target item was positioned either to the left or right of the participant, amongst 29 other items of varying colour. In the present study, there was only one item on either side of the screen. Previous evidence has found that if the stimuli are large enough (in terms of the proportion of visual field they inhabit), colour can be detected and distinguished parafoveally (Nagy, Sanchez & Hughes, 1990), meaning participants may have been able to identify the side of the screen cued quickly using peripheral vision. If this were the case, then there would have been no advantage of using a gaze cue with a featural language cue compared to a spatial language cue, as there is effectively information overlap in either case.

The analysis of congruity effects allowed for the investigation of what happens when gaze and language cue different targets, and how this is affected by reliability. Study 2 provided evidence that in the real world gaze cues were ignored when incongruent to language, however, language cues were always correct in that task, while gaze cues were either supportive or distracting. In the present study, gaze and language cues were equivalent. Although this equivalence may not be ecologically valid, it did allow for any innate preference for either cue to be identified. Starting with equally reliable cues, no bias was found for either language or gaze when both cues were 50% reliable. Separating analysis by reference type again showed no sign of a preference for either cue. However, the results from this session are perhaps ultimately uninformative because participants all self reported losing motivation. This session was unique in this experiment, as it was the only session in which participants could not employ a strategy to perform significantly better than chance. Because the participants reported that they disengaged with the task, we cannot infer anything from these results about any cue preferences when both cues were equally unreliable.

A better way to investigate whether participants showed a natural preference for either cue is to investigate the session in which the cues were equally reliable and also more reliable than chance (80%). Here, participants followed the language cue more often when the cues were incongruent, suggesting that language is the dominant cue. This finding supports the hypothesis, as well as the results of earlier studies. Macdonald and Tatler (2013) showed that when language and gaze cues were both informative participants used language and ignored gaze, while Knoeferle and Kreysa (2012) found that when processing difficult sentences, participants ignored supportive gaze cues. Both studies suggest a preference for language cues over gaze cues, however it is arguable that in both paradigms the cues were not given equal prominence; in Macdonald and Tatler (2013) gaze cues had to be actively sought out, while verbal instructions were directed toward the participant and in Knoeferle and Kreysa (2012) language comprehension was central to the task and gaze cues were only supportive. In the present study, however, the paradigm gave equal prominence to language and gaze cues, both explicitly (by telling participants that verbal and non-verbal cues were there to help them) and implicitly (by making the reliability of both cues 80%). Language cues have therefore been shown to be preferred over gaze cues when both cues provide the same quality of information.

It is worth considering whether this preference for language may be peculiar to this paradigm rather than generalisable to real world interactions. In one way, the preference for language over gaze seems intuitive, as language is our primary mode of communication. Anecdotally, however, it is easy to think of a situation in which someone says “left” when they actually mean “right”. It is harder to think of a situation where someone accidentally looks in the opposite direction to where they intended to cue another’s attention. In this case, it seems intuitive that gaze cues should be trusted over language when there is any incongruity between the two cues. It may be the case that gaze cues are indeed favoured over language in certain natural situations, but that the nature of the cue stimuli in this study biased participants to language cues. The language cues in this paradigm were intentionally embedded within natural instructional sentences in order to keep the cues as natural as possible and maximise similarity with the language cues used in Study 2. It could be argued that this design may have had the side effect of leading to a preference for language cues, as the sentences explicitly directed participants to a target, whereas the gaze cues required participants to make an inference about where the cueing person was attending. However, the informativeness of gaze cues could be learned very quickly in this paradigm and so these cues essentially became explicit in their meaning. For this reason, it is unlikely that the preference for language when both cues were equal was just due to the explicit nature of the instructional sentences.

Despite reference type having no significant effect on the main dependent variables in the analysis of all sessions, there was an approaching significant effect of reference type on cue preference when cues were equally reliable. In the session with 80% gaze and language reliabilities, when there was an incongruity between gaze and language, participants were more likely to choose language over gaze for both spatial and featural instructions. However, the proportion of trials in which the language cue was followed was higher for trials with featural instructions compared to spatial instructions. This shows that the preference for language cues was decreased when these cues were spatial in nature. It is not entirely clear why there is this difference. To speculate, the direct incongruity of spatial information provided by language and gaze may have slightly decreased the tendency for participants to follow language when cues were incongruent. It may be that a spatial gaze cue interferes more strongly with a conflicting spatial language cue than a conflicting featural language cue because the spatial language provides the same form of information as the gaze cue. Following the featural cue may have merely required selecting a colour stimulus upon hearing the name of a colour, which may have been less susceptible to distraction from a conflicting visual-spatial cue.

When the language cues were 100% reliable and gaze cues less reliable, there was an interaction between congruity and gaze cue reliability. Although performance was very good in all conditions in these sessions, incongruent gaze cues were more detrimental to accuracy when gaze was 80% reliable compared to when it was 50% reliable. To discount the possibility that this effect was driven by the lower number of incongruent trials in the 80% gaze reliability session pulling down the overall proportion, this analysis was repeated on the first and final five blocks of the 80% gaze reliability session and the first and final two blocks of the 50% gaze reliability session. The same pattern of results was shown in these comparisons, discounting the possibility that the interaction was caused by the difference in overall number of incongruent trials between gaze reliability conditions. These results suggest the participants were more likely to follow the gaze cue stimuli when they were more reliable, even though the optimal strategy was to ignore gaze cues and follow language cues. The results from the accuracy analysis were supported by the finding that gaze stimuli had a larger facilitative effect (when congruent) and detrimental effect (when incongruent) on the time to first fixate the target and response time when they were 80% reliable compared to 50% reliable. Additionally, although there was no difference in the beneficial effect of congruent gaze cues between gaze cue reliability conditions in the correct initial saccade measure, there was an increased detrimental effect of incongruent cues when gaze cue reliability was 80% compared to 50%. Combined, these findings suggest that 1) gaze cues are not completely ignored, even when language is 100% accurate and 2) the extent to which gaze cues are ignored is inversely related to their perceived reliability. This shows, like the accuracy analysis, that the gaze cues had more influence when they were more reliable; reliable cues slowed down performance when incongruent with language and sped up performance when congruent with language.

Overall, the distracting gaze cue results are in line with the hypothesis that the gaze cueing effect would be persistent, but influenced by reliability. When language was 100% accurate, gaze need not (and should not) have been used, however the gaze cues still slowed down and disrupted performance. These results are typical of a gaze cueing paradigm study (Friesen & Kingstone, 1998). However, gaze cues have also been found to have less influence when they are less reliable, supporting previous findings that the attentional effects of gaze cues are affected by reliability (Hill et al, 2010). Using a real world interaction, in Chapter 3 evidence was reported showing that gaze cues were sought out less when accompanied by language that provided overlapping semantic (spatial) information. This is because in cases where language did not provide overlapping semantic information (featural condition), the gaze cues were more exclusively informative. However, in the present study we have seen that when gaze cues were centrally presented on a screen (i.e. gaze seeking is out of the participant's control), gaze following was not inhibited completely by more informative language, but instead gaze following was modulated by the reliability of the cues. This is likely due to the centrally presented gaze cue stimuli in the present study; the selective gaze utilisation behaviour found in Study 2 may have been manifest in the gaze seeking stage of gaze utilisation, a stage that is not present in this lab-based paradigm.

The same effects were not found when language reliability varied alongside 100% reliable gaze cues. The time to first fixate the target was significantly slower when language reliability was 50% compared to 80%. Furthermore, participants were significantly slower to launch their first saccade in the 50% language reliability sessions compared to the 80% reliability sessions. One possibility for these findings is that they are due to the high rate of incorrect language cues interfering with participants' performance in the task. If this was the case, then this finding that unreliable language caused an overall detriment to performance, whereas unreliable gaze did not, suggests that gaze cues may be easier to ignore than language cues, or at least that the distracting effect of unhelpful language cues lasted longer. This may be indicative of the relative importance of language compared to gaze in natural communication. These findings, combined with the preference shown for language cues when presented alongside equally reliable gaze cues, supports the hypothesis that language is the dominant cue. However, as with the language preference found in the 80% gaze and language reliability condition, it is possible that this finding is peculiar to this paradigm, as the language cues are more explicitly instructional than the gaze cues.

There was no overall effect of language congruity for the time to first fixate the target or response time, suggesting that participants followed the more reliable gaze cues and were uninfluenced by instances of incongruent language. It is surprising to find that the congruity of a less reliable gaze cue effects performance, but that the same effects are not found with language cues, particularly given the evidence shown that language is the preferred cue. The lack of a congruity effect is likely due to the gaze cues being responded to faster than language cues and so when the former were entirely reliable, the latter had less time to disrupt processing (response times were faster for 100% gaze reliability conditions than 100% language reliability conditions). The main contributing factor to the increased time to respond to language was the time of cue disambiguation. Language cues did not disambiguate the target until the final word was heard, whereas it is arguable that the gaze cue was unambiguous as soon as the head began to turn. This explanation is supported by the accuracy results for these sessions, which show a clear congruity effect, despite the lack of an effect on the time measures. The response time measure necessarily only includes correct trials and time to first fixate the target measure only includes trials in which the target was looked at. The congruity effect for accuracy is likely due to the incongruent trials in which participants made a selection after hearing and processing the language cue. These slow trials were more likely not to be included in the time to first fixate the target and response analysis as they either didn’t include a fixation on the target or were incorrect.

This experiment investigated the effect of varying gaze and language cue reliability on attention. Participants strategically made use of the most reliable cue to complete the task and favoured language when reliabilities were equal. The results show that changing the reliability of language affects attention differently than changing the reliability of gaze. Language appears to be the favoured cue and causes overall disruption when unreliable, whereas gaze is disruptive when incongruent to more informative language, but more so when gaze is more reliable than chance.

Although the relative timings of the gaze and language cues matched those used in Study 2, as well as matching the way the cues have been shown to correspond in natural language production (Hanna & Brennan, 2007), these timings were problematic for two reasons. Firstly, embedding the language cues within natural and explicit instructional sentences may have led to an artificial preference for these cues over the gaze cues. Secondly, the naturalistic approach that was used moved the onset of the disambiguating language cue to after the onset of the gaze cue, so the seemingly distinct effects of distracting language and gaze cues may be the result of the onset times of these cues, rather than the inherent properties of these particular cues. The next study in this thesis sacrifices the naturalistic sentences in order to make the gaze and language cues as equivalent as possible. Additionally, the timings of the onset of both of the cues are synchronised in order to reduce the possibility of one cue being processed before the other.

Chapter 5:

Varying the reliability of gaze and language (ii) – Studies 4 and 5

Introduction

Study 2 provided evidence showing that the type of verbal reference used in an instruction can affect the extent to which listeners use supportive gaze cues; verbal references that increased the relative informativeness of the gaze cues resulted in increased gaze utilisation. Study 3 investigated the relationship between the informativeness of gaze and language in a more controlled environment and found that altering the reliability of a gaze cue affected how distracted participants were from a concurrent language cue. The same pattern was not found when altering the reliability of language cues alongside a concurrent gaze cue. These results suggest that altering the reliability of a gaze cue has a distinct effect from altering the reliability of a language cue. However, it is possible that some of these differences have arisen from the nature of the paradigm used. Firstly, the language cues used in Study 3 were embedded within natural and explicit instructional sentences, which may have led to participants favouring them over the gaze cue stimuli. Secondly, the onset of the crucial word in the language cues used in Study 3 always occurred after the onset of the gaze cue, so it may be the case that the language cues simply did not have enough time to affect performance.

The studies outlined in the present chapter aim to further investigate the effects of altering concurrent gaze and language cues, as well as assessing these stimuli by comparing them in isolation, by simplifying the language cueing stimuli of Study 3. In Study 4, the language cues were simplified to cueing words, rather than explicit instructional statements in order to increase the equivalence of the gaze and language stimuli. This simplification of the language stimuli also allowed for the onset time of the gaze cue and the onset time of the cueing word to be identical in Study 4, thus reducing the chance that any effects were driven by the timings of the cues.

With the removal of the natural instructional context of the language cues and the synchronisation of the cue onset times, the equivalence of the gaze and language cues can be tested. If participants treated these cues equivalently, then subordinate gaze cue effects should match those found in Study 3, but congruity effects should also be found with subordinate language cues as, in contrast to Study 3, the onset time of the language cues was identical to the gaze cues. Any differences found between the effects of subordinate gaze cues and subordinate language cues may indicate a preference for one of the cues over the other.

Study 4

Method

Participants. Ten students from the University of Dundee took part in four one-hour sessions over a five-day period. They each received £10 for their participation.

Materials and eye movement recording. Language cues were made up of a two-second initial audio clip (“Select”), one of two two-second spatial-determiner audio clips (“Left” or “Right”) or one of twelve two-second featural-determiner audio clips (“[Name of colour]”). The end of “Select” in the initial audio clip occurred 1,700 ms into the clip and was followed by 300 ms of silence. Each of these initial audio clips began with 1,170 ms of silence, until the onset of the word “Select”. The determiner audio clips began immediately with the onset of the object name and the remainder of the two-second clips was filled with silence. The exact duration of these silent sections varied depending on the duration of the particular descriptor word. Picture stimuli, gaze cue stimuli and all other materials and eye-tracking equipment were identical to those used in Study 3. The mean calibration error was .416° (SD = .128°) of visual angle. As with Study 3, at the beginning of each trial participants fixated a marker in the middle of the screen. The average error for the experiment was .530° (SD = .295°) of visual angle.

Design. The present experiment used a within-subjects design. Between sessions there were two independent variables: The probability that the gaze cue was correct (50%, 80%, 100%) and the probability that the language cue was correct (50%, 80%, 100%). In each of the four sessions, one of the cues was always 100% reliable and the other was either 80% or 50% reliable. As with Study 3, these four sessions were used to investigate whether there was a detrimental effect of incongruent gaze and language cues and identify whether these effects were mediated by subordinate cue reliability. Within each session, there were 16 blocks of trials and in each of these the proportion of correct cues matched the proportion for the whole session. Within each block whether or not each cue was correct was varied (except when the cue was 100% reliable). Half of the language cues in each block used spatial references (left or right) and the other half used featural references (colour).

Procedure. The procedure for this experiment (Figure 5.1) differed from that of Study 3 in only two ways. Firstly, each participant carried out only four sessions instead of nine. These were the four sessions in which one cue was 100% reliable and the other cue less reliable. Secondly, instead of embedding the language cues in full sentences, simple one word audio files were used (See *Materials and eye movement recording* section). The initial audio file ended at the two-second point in the video, which is when the determiner audio file began playing.

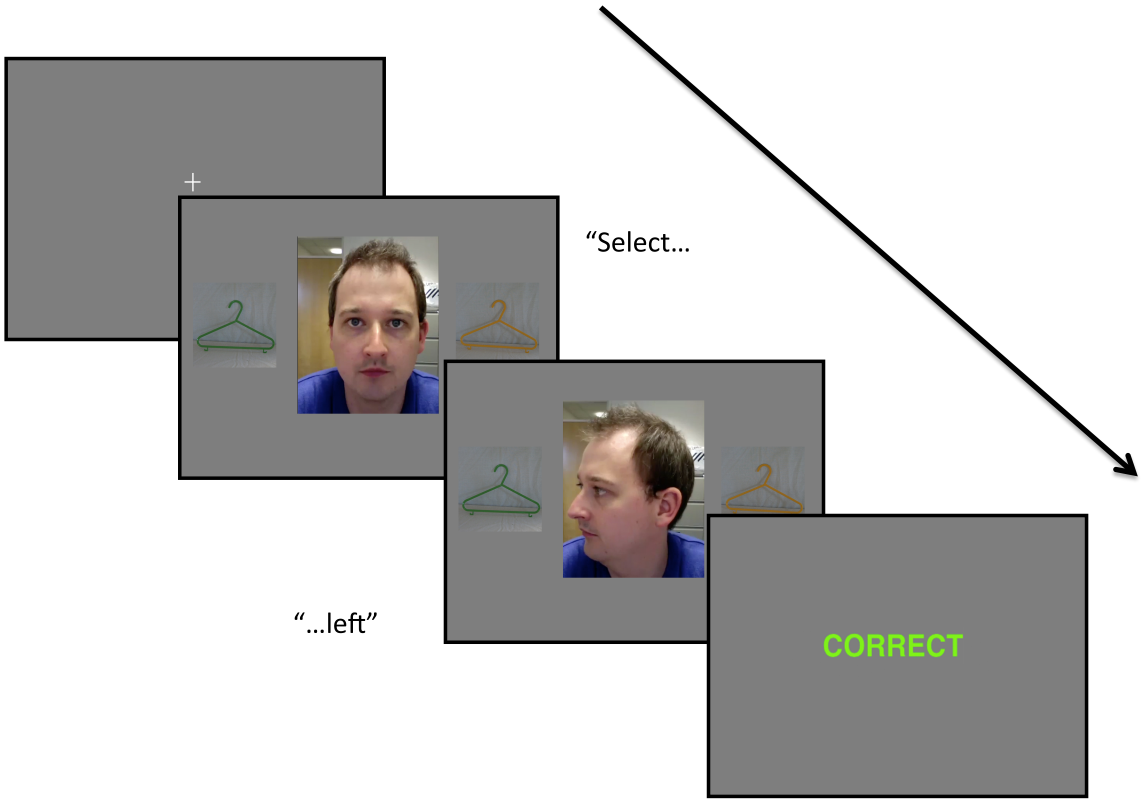

Figure 5.1. Outline of a trial. After the central fixation point was fixated, a video began showing a face looking at the participant, with pictures of objects on either side that differed only in colour. While this video played, the word "select" could be heard. After two seconds the head began to turn towards one of the targets while the determiner word was heard. After making a decision the participant received immediate feedback

Analysis. The same five main dependent variables used in Study 3 were used here: two performance measures (accuracy and response time) and three eye movement measures (first saccade direction, first saccade latency and time to first fixate the target). In the present study, rather than all nine possible sessions, participants only took part in sessions in which one cue was 100% reliable and the other cue less reliable. These sessions were analysed as outlined in the *Analysis* section of the Study 3 *Method*.

Results and Discussion

How participants behaved when faced with two opposing cues was central to the analyses in this study. Specifically, the focus was on whether the reliability of a less-reliable cue affected participant responses, even though the other cue was completely reliable. In the four sessions used in this study, as with Study 3, there were two levels of congruity: 1) cues congruent with each other and correct and 2) cues incongruent with each other, with the less reliable cue incorrect. In Study 3, the effects of less reliable gaze cues on accuracy, first saccade direction and latency, time to first fixate the target and response time were different from the effects of less reliable language cues. It is possible that these differences found in Study 3 were the result of the difference in onset times of the cue stimuli. Alternatively, these differences may have arisen as a result of the language cues being more explicitly informative, as they instructed participants how to respond using natural spoken sentences. In this study, the language cues were simplified and gaze and language onset times were identical.

Accuracy. Figures 5.2a and 5.2b show the accuracy results for both congruent and incongruent trials in all four sessions. An LMM (Model 5.1) of accuracy (using congruity and gaze reliability as fixed factors and subject and item as random factors) for trials where language reliability was 100% (Figure 5.2a) showed that participants were significantly more accurate in congruent trials, $\beta = 2.352$, $SE = .350$, $z = 6.721$, $p < .001$. There was no significant interaction between congruity and gaze reliability, $\beta = .782$, $SE = .528$, $z = 1.480$, $p = .139$, nor was there any overall effect of gaze reliability, $\beta = -.297$, $SE = .428$, $z = -.692$, $p = .489$, on accuracy.

The significant congruity effect and the lack of an overall gaze reliability effect are in line with the findings from Study 3. However, the interaction between congruity and gaze reliability found in Study 3 (Chapter 4, Figure 4.10a) was not found here. This interaction was not significant because, in contrast to the findings of Study 3, a Tukey test revealed a significant difference between accuracy in congruent and incongruent trials in the session with 50% gaze reliability ($p < .001$). Although, this difference was not significant in Study 3, the difference was in the same direction, so it is possible that with five participants the experiment lacked the power to produce a significant result. Alternatively, it may be that the language cues in Study 3 were less susceptible to distraction from unreliable gaze cues as they were embedded in a naturalistic instruction sentence, whereas in the present study the language cues were less natural and more simplistic. Regardless of the reason, the performance in incongruent trials was still significantly worse in the 80% gaze reliability session compared to the 50% reliability session ($p < .001$). This result supports the findings from Study 3 and provides further evidence that subordinate gaze cues are followed more as their reliability increases.

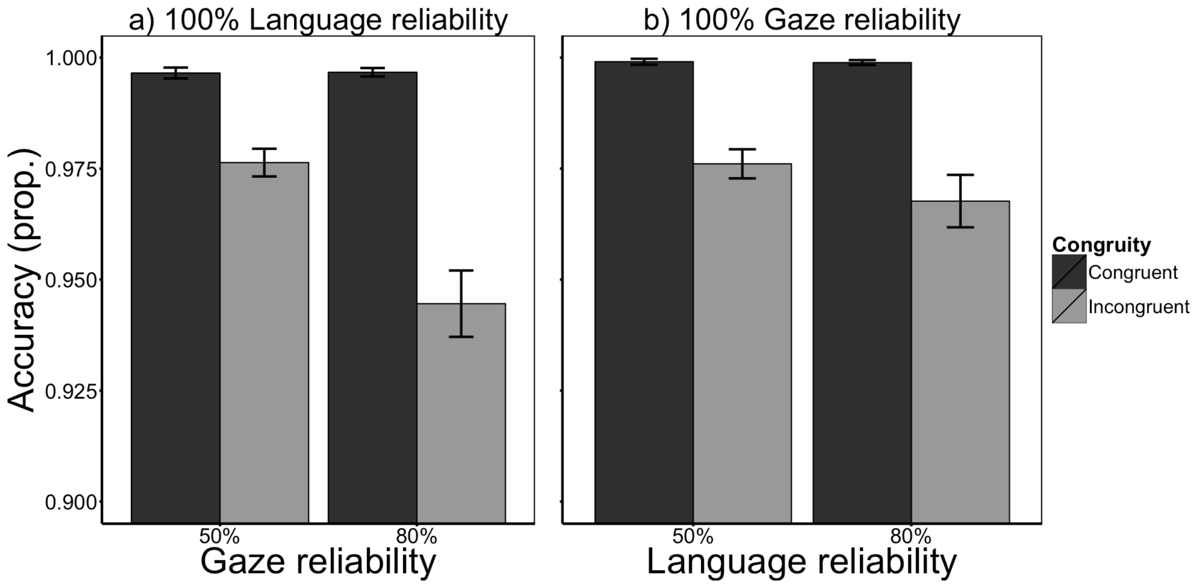

Figure 5.2. The mean accuracy (proportion) for participants in congruent and incongruent cue trials for a) sessions with 100% language reliability and 50% and 80% gaze reliabilities and b) sessions with 100% gaze reliability and 50% and 80% language reliabilities.

An LMM (Model 5.2) (using congruity and gaze reliability as fixed factors and subject and item as random factors) of accuracy for trials in which gaze reliability was 100% (Figure 5.2b) showed that participants were significantly more accurate in congruent trials than incongruent trials, $\beta = 3.305$, $SE = .470$, $z = 7.032$, $p < .001$. Language reliability did not have an overall effect, $\beta = -.326$, $SE = .653$, $z = -.500$, $p = .617$, nor was there a significant interaction between the two factors, $\beta = .176$, $SE = .905$ $z = .194$, $p = .846$.

The results from the LMM above show the same significant effect of congruity that was found in Study 3 (Chapter 4, Figure 4.10b). This finding supports the findings in

Study 3 that showed that incongruent language was detrimental to accuracy when gaze was 100% reliable whether the reliability of language was 50% or 80%.

First saccade direction. The proportion of first saccades after the onset of the cues that were in the direction of the target is shown in Figures 5.3a and 5.3b. For the two sessions with 100% reliable language cues and less reliable gaze cues, an LMM (Model 5.3) (using congruity and gaze reliability as fixed factors and subject and item as random factors) showed no significant effect of gaze reliability, $\beta = .059$, $SE = .108$, $z = .550$, $p = .583$, but a significant effect of congruity, $\beta = 2.318$, $SE = .454$, $z = 5.105$, $p = 001$, and a significant interaction, $\beta = .935$, $SE = .116$, $z = 8.067$, $p < 001$ (Figure 5.3a). Pairwise comparisons showed that the proportion of initial saccades in the correct direction for congruent trials was significantly higher when gaze cue reliability was 80% compared to 50% ($p < .001$), but there was a significantly higher proportion of correct first saccades in the 50% gaze reliability condition when gaze cues were incongruent ($p < .001$).

These results are broadly supportive of the first saccade direction results of Study 3 (Chapter 4, Figure 4.12a). One difference, however, was that in the present study, there was a significant difference between congruent trials between 50% and 80% gaze reliabilities, which was not present in Study 3. It is worth noting however, that there was a non-significant difference in the same direction in Study 3, and the difference between the 50% and 80% gaze reliability trials in the incongruent condition was larger than the difference in congruent trials in the present study and Study 3. It may be that the statistical power of the present study allowed for this difference to reach significance.

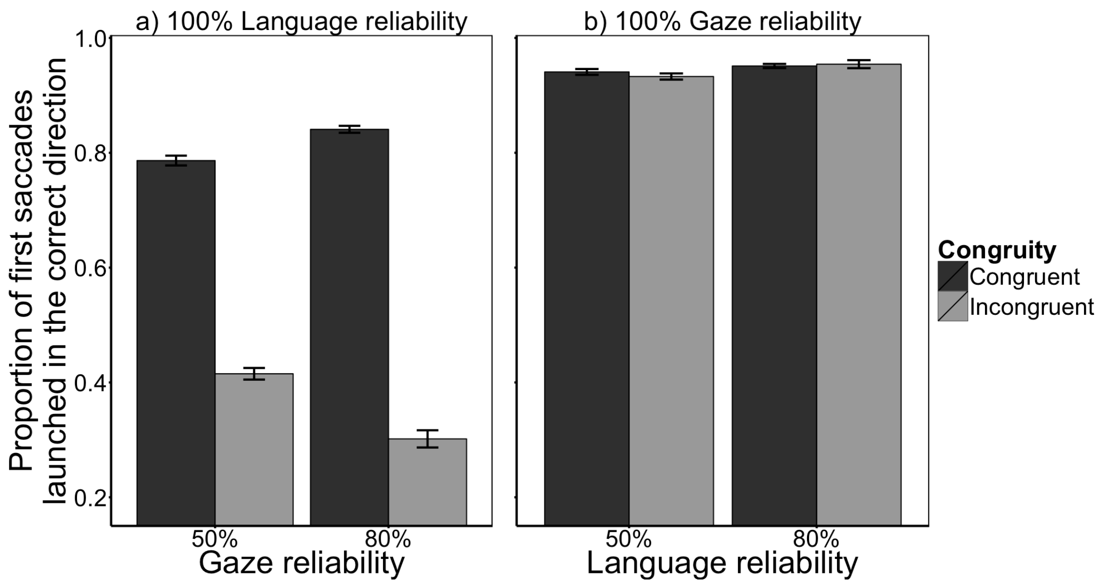

Figure 5.3. The proportion of first saccades launched in the correct direction after the onset of the cues for participants in congruent and incongruent cue trials for a) sessions with 100% language reliability and 50% and 80% gaze reliabilities and b) sessions with 100% gaze reliability and 50% and 80% language reliabilities

An LMM (Model 5.4) of proportion of correct first saccades for trials in which gaze reliability was 100% (Figure 5.3b) showed no overall difference between the proportion of initial saccades launched in the correct direction in congruent and incongruent trials, β = .017, *SE* = .135, z= .126, *p* = .899. Likewise, there was no significant difference between sessions in which language reliability was 80% compared to 50%, β = .146, *SE* = .349, *z* = .418, *p* = .676, nor was there any interaction between these two factors, β = -.230, *SE* = .231, *z* = -.996, *p* = .319. These findings were in line with the equivalent findings from Study 3 (Chapter 4, Figure 4.12b), showing the language reliability had no effect on first saccade direction when gaze was 100% accurate, even when the onset time of both of these cues was identical.

First saccade latency. LMMs of first saccade latency were carried out for all sessions (Figures 5.4a and 5.4b). An LMM (Model 5.5) of saccade latency (using congruity and gaze reliability as fixed factors and subject and item as random factors) for trials in which language reliability was 100% (Figure 5.4a) showed no overall difference between the mean saccade latency in congruent and incongruent trials, $\beta = -4.832$, $SE = 3.616$, t= -1.336, $p = .178$, nor a significant difference between sessions in which gaze reliability was 80% compared to 50%, $\beta = -6.842$, $SE = 10.277$, $t = -.666$, $p = .487$. No interaction was found between these two factors, $\beta = 6.815$, $SE = 4.855$, $t = -1.404$, $p = .163$. These results were in line with the equivalent findings of Study 3 (Chapter 4, Figure 4.13a) that showed no effect of either gaze reliability or congruity on first saccade latency when language cues were 100% reliable.

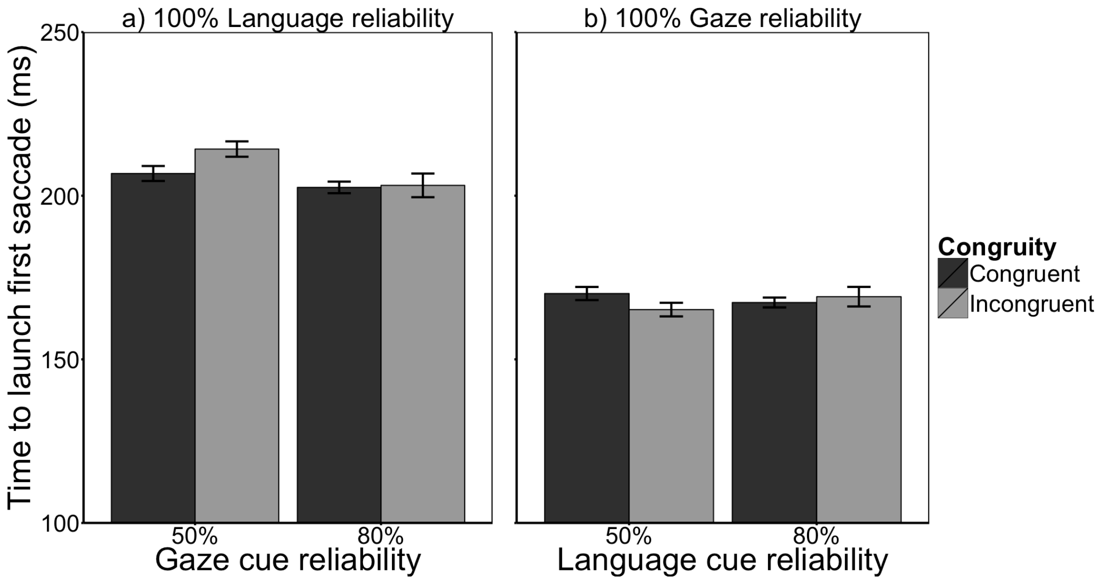

Figure 5.4. The mean first saccade latency (ms) for participants in congruent and incongruent cue trials for a) sessions with 100% language reliability and 50% and 80% gaze reliabilities and b) sessions with 100% gaze reliability and 50% and 80% language reliability

An LMM (Model 5.6) of first saccade latency for trials in which gaze reliability was 100% (Figure 5.4b) showed no overall difference between the mean saccade latency in congruent and incongruent trials, β = 1.378, *SE* = 2.245, *t* = -.614, *p* = .539, nor was there an effect of language reliability, β = 2.147, *SE* = 8.676, *t* = .247, *p* = .794. There was also no significant interaction between these two factors, β = -7.135, *SE* = 4.769, *t* = .-1.496, *p* = .133.

The present results differ from those of Study 3 (Chapter 4, Figure 4.13b). In Study 3, first saccade latency time was significantly longer for the 50% language reliability session regardless of congruity. This supported the time to first fixate target and response time results in Study 3 (Chapter 4, Figures 4.14b and 4.15b respectively) that showed time to first fixate the target and response times were slower when language was 50% reliable. In the present study, as with time to first fixate target and response time, there was no significant effect of gaze reliability on first saccade latency. This results provides further support that the overall detrimental effect found in Study 3 for unreliable language was not present in this study, where the language cues were presented outwith the context of a natural instructional sentence.

Time to first fixate the target (FT). Figures 5.5a and 5.5b show the time to first fixate the target for congruent and incongruent trials in the four sessions. An LMM (Model 5.7) of FT (using congruity and gaze reliability as fixed factors and subject and item as random factors) for trials where language reliability was 100% (Figure 5.5a) showed participants to be significantly quicker to fixate the target in congruent trials, $\beta = -142.05$, $SE = 22.93$, $t = -6.195$, $p < 001$, as well as a significant interaction between congruity and gaze reliability, $\beta = 45.35$, $SE = 16.987$, $t = -2.670$, $p = .015$. There was no overall effect of gaze reliability, $\beta = -6.767$, $SE = 10.634$, $t = -.636$, $p = .511$. Post-hoc Tukey tests showed significant differences in time to first fixate the target between the gaze cue reliability conditions for both congruent and incongruent cues. However, the time to first fixate the target was shorter in the 80% reliable condition, relative to the 50% reliable condition, for congruent cues, $p < .001$, but longer for incongruent cues, $p < .001$.

The LMM results above are in line with those found in Study 3 (Chapter 4, Figure 4.14a). Participants took significantly longer to fixate the target when gaze cues were incongruent to 100% reliable language cues, but this distracting effect was greater when gaze cues were 80% reliable compared to 50% reliable. Furthermore, when gaze cues were congruent, participants were quicker to hit the target when gaze cues were 80% reliable compared to 50% reliable. These results support Study 3 and provide further evidence that gaze cues had more influence on performance when they were more reliable, even when accompanied by perfectly reliable language cues.

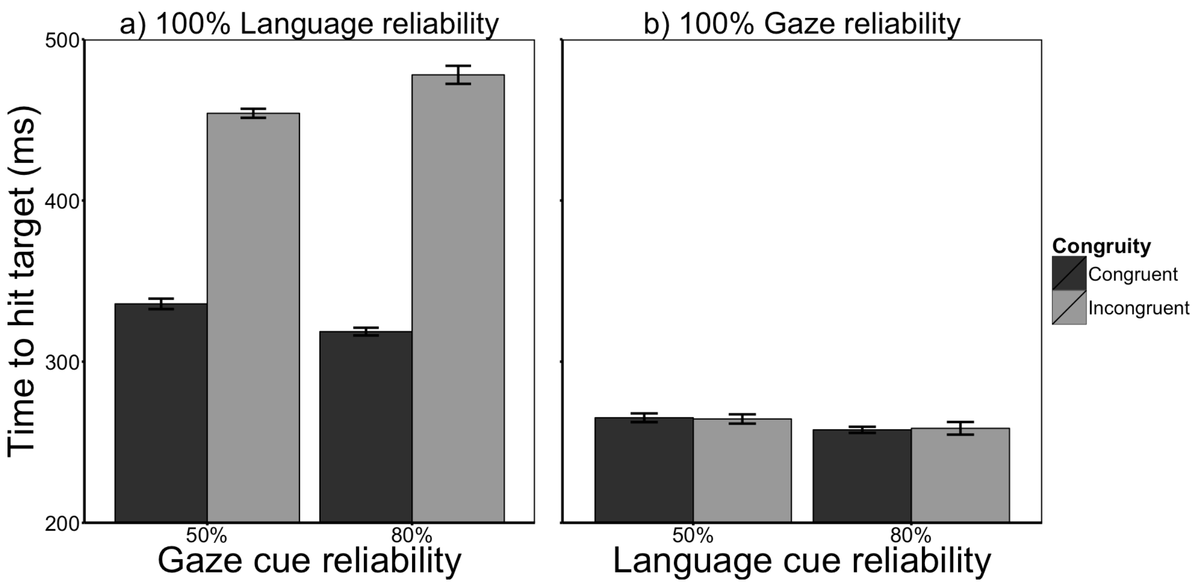

Figure 5.5. The mean time to fixate target (ms) for participants in congruent and incongruent cue trials for a) sessions with 100% language reliability and 50% and 80% gaze reliabilities and b) sessions with 100% gaze reliability and 50% and 80% language reliabilities

An LMM (Model 5.8) of time to first fixate the target for trials in which gaze reliability was 100% (Figure 5.5b) showed no overall difference between the time to first fixate the target in congruent and incongruent trials, β = -.820, *SE* = 3.216, *t* = -.255, *p* = .784, no effect of language reliability, β = -5.078, *SE* = 12.213, *t* = -.416, *p* =

.666, nor any interaction between these two factors, β = -1.283, SE = 6.728, t = -.191, p = .838.

The above results differed from those found in Study 3 (Chapter 4, Figure 4.14b), but not as anticipated. It was hypothesised that if language and gaze cues were treated equivalently by participants, a congruity effect, similar to that shown in Figure 5.5a would be found, as in each trial the onset time of the language cue was identical to the onset time of the gaze cue. However, no congruity effect was found. One reason for this may be that participants were more resistant to distracting verbal cues than distracting gaze cues, therefore the incorrect language cues had no effect on the time taken to fixate the target cued correctly by gaze. However, a more likely explanation, particularly given that participants were quicker to fixate the correct target in the 100% gaze reliability sessions overall, is that the processing time of the audio linguistic cue in this paradigm was longer than the processing time for the visual gaze cue. This increased time would mean that participants might not have been distracted by incongruent language cues until after they had already planned a saccade to the gaze cued-at location. The difference found between the results in the present study and Study 3, was that there was no overall difference between performance in the 50% language reliability session and the 80% language reliability session. In Chapter 4, it was argued that hearing regular incorrect sentences in the 50% reliable language condition led to a slow down in overall performance. This explanation is challenged by the lack of a language reliability effect in the present experiment. It is possible, however, that the language cues in the present study had no overall effect because they were not embedded within a natural sentence. It may be that hearing incorrect instructional cues in the context of natural sentences provides a distracting effect that is not shared with isolated cueing words.

Response Time (RT). An LMM (Model 5.9) of RT (using congruity and gaze reliability as fixed factors and subject and item as random factors) for trials in which language cues were 100% reliable (Figure 5.6a) showed congruent trials to have significantly quicker response times, β = -60.785, *SE* = 14.000, *t* = -4.342, *p* < .001, and a significant interaction was found between congruity and gaze reliability, β = -43.329, *SE* = 15.390, *t* = -2.815, *p* = .008. There was no overall effect of gaze reliability, β = 2.294, *SE* = 26.722, *t* = .086, *p* = .928. Post-hoc Tukey tests showed that that incongruent trials had a significantly slower RT when gaze cue reliability was 80% compared to 50%, *p* < .001, and that congruent trials had a significantly slower RT when gaze cue reliability was 50% compared to 80% *p* = .001.

As with the time to first fixate the target results for these sessions, the LMM results for response time are in line with those found in Study 3 (Chapter 4, Figure 4.15a). When 80% reliable, gaze cues were found to have a larger detrimental effect when incongruent and a larger beneficial effect when congruent relative to when gaze was 50% reliable. This provides more evidence that gaze cues are used more and affect task performance more when they are perceived to be more reliable.

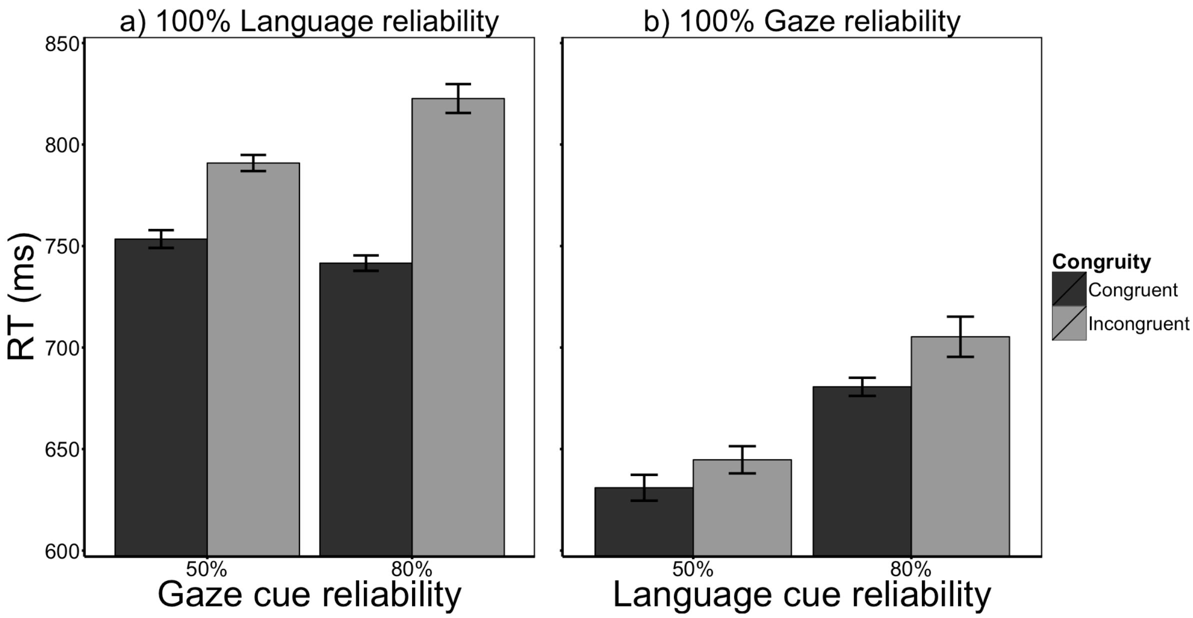

Figure 5.6. The mean response time (ms) for participants in congruent and incongruent cue trials for a) sessions with 100% language reliability and 50% and 80% gaze reliabilities and b) sessions with 100% gaze reliability and 50% and 80% language

When gaze reliability was 100% (Figure 5.6b) there was a significant congruity effect (Model 5.10), β = -20.470, *SE* = 9.515, t = -2.151, p = .042. There was no significant effect of language reliability, β = 63.169, *SE* = 42.807, t = 1.476, p = .141, nor any interaction between language reliability and congruity, β = -12.040, *SE* = 11.408, t = -1.055, p = .294.

These results were very different from the equivalent findings in Study 3 (Chapter 4, Figure 4.15b). The present results showed a clear congruity effect, with participants slower to respond when the language cues were incongruent with the correct gaze cues. This effect may have been present because the language and gaze cues shared an onset time, whereas there was a delay in onset of the crucial word in Study 3.

However, this congruity effect was only present in the response time results but not the results for time to first fixate the target. It appears that in the time taken for participants to respond to a 100% reliable gaze cue by planning a saccade in the cued direction, incongruent language cues did not cause disruption, however by the time participants made a decision as to which trigger to pull, incongruent language cues did cause disruption, suggesting that although the cues were presented at the same time, the gaze cues were processed earlier. This argument is supported by the overall shorter time to first hit the target and response time in the 100% gaze reliability sessions compared to the 100% language reliability sessions.

Taking all of the findings of the present study together leads to the suggestion that the gaze stimuli were processed earlier than the language stimuli, despite the synchronisation of cue onset times. Specifically, it was found that subordinate incongruent language cues had no effect on the early trial measures (first saccade latency, first saccade direction and time to fixate target), but a clear detrimental effect on the late trial measure of response time, whereas subordinate incongruent gaze cues had a detrimental effect on all measures. To further investigate the potential difference in processing time between the gaze and language cues, the five dependent variables used in the present study were measured for each cue in isolation in Study 5.

In Study 2, evidence was found suggesting that incongruent gaze cues were more distracting to accompanying spatial language cues than featural language cues. No such effect was found in Study 3. It is possible that the difference in processing time for the language and gaze cues in Study 3 and 4 led to there being no effect; the gaze cues may have been used before the linguistic reference type had been determined. By isolating gaze and language cues in Study 5, it was possible to observe switching effects. Task switching effects are a common tool used in cognitive science to investigate executive function (Kiesel et al., 2010). The costs of switching have been found to be largely influenced by the tasks that are being switched between (Kiesel et al., 2010; Koch, Gade, Schuch & Philipp, 2010). Switching from a spatial gaze cue to a spatial language cue may cause less disruption than switching to a featural language cue, as the reference type of the cue changes. Alternatively, switching to the same type of reference in a different form may cause disruption analogous to the findings in Study 2, with inherently spatial gaze cues causing disruption to the interpretation of spatial language cues.

Study 5

Method

Participants. Thirteen undergraduate students from the University of Dundee took part in one hour-long session. They all received course credit for their participation

Materials and eye movement recording. Materials and eye movement recording procedures were identical to those used in Study 4, except for the addition of 16 extra four-second videos. These videos were composed of the first two-seconds of the videos used in Studies 3 and 4, played normally and then in reverse, to make four-second videos of a face staring forward. The mean calibration error was .393° (SD =

.109°) of visual angle. The average error of pre-trial marker fixations was .456° (SD = .285°) of visual angle.

Design. The present experiment used a within-subjects design. Unlike Studies 3 and 4 there was only one session, and all cues were always reliable. Within this session there were again 480 trials in 16 blocks. Within each block the type of cue used varied (gaze or language). Half of the trials featured a language cue and the other half featured a gaze cue. There was only one cue per trial.

Procedure. In contrast to Studies 3 and 4, the present experiment involved only one hour-long session. In this session all of the cues were 100% reliable and the participants were informed that they would need to follow these cues to complete the task. However, each trial only contained one cue: half of the trials had a language cue and the other half had a gaze cue. Trials were mixed, rather than blocked by cueing condition. This allowed for the investigation of switching effects. In the language cue trials, participants were presented with the same audio stimuli used in Study 4 but different visual stimuli. Instead of a head turning after 2-seconds, they were shown a video of a face staring forward for the duration of the trial (Figure 5.7a). In the gaze cue trials participants saw the head-turning stimuli, but were provided with no audio stimuli (Figure 5.7b).

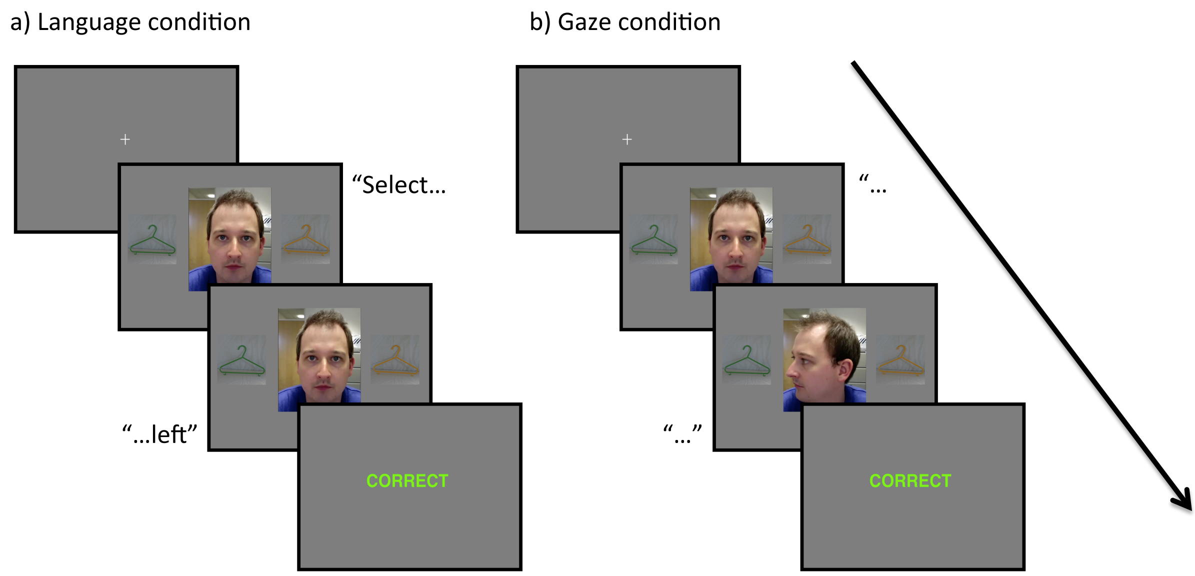

Figure 5.7. Examples of a) language and b) gaze condition trials. In the language condition, a video began showing a face looking at the participant, with pictures on either side that differed in colour. While this video played the word "select" could be heard. After two seconds the determiner word was heard, while the face continued to look at the participant. In the gaze condition, no words were heard, but after two seconds the head turned to the left or right. After making a decision the participant received immediate feedback

Analysis. The same five main dependent variables used in Study 3 and 4 were used here: two performance measures (accuracy and response time) and three eye movement measures (first saccade direction, first saccade latency and time to first fixate the target). As well as comparing different cue types, different reference types were compared within the language cues. In Study 2, a clear effect of reference type was found in the gaze seeking and following measures, however no effect of reference type was found in Study 3 when the gaze seeking stage of gaze utilisation was removed. In the present study, the processing time of featural and spatial cues were compared, as well as the effects of switching to either type of language cue from a gaze cue. It is possible that switching from a spatial gaze cue to a spatial language cue caused less disruption than switching to a featural language cue, as the reference type as well as the form of the cue changed. The difference in the dependent variables between language and gaze trials was investigated using linear mixed effects models (LMMs), which were carried out as outlined in previous chapters.

Results and Discussion

There is evidence from Study 4 that although gaze and language cues were used concurrently, gaze cues were responded to earlier. To investigate this, the gaze and language cues were used in isolation and accuracy, first saccade direction, first saccade latency, time to first fixate the target (FT) and response time (RT) were measured.

Accuracy. The proportion of correct trials is shown in Figure 5.8a. Although accuracy was very high in both conditions, an LMM (Model 5.11), using type of cue as a fixed factor and object and subject as random factors, showed a significant difference across cue condition, $\beta = -3.328$, $SE = 1.216$, $z = -2.736$, $p < .006$. This difference may be due to participants mishearing the language cueing word on some trials.

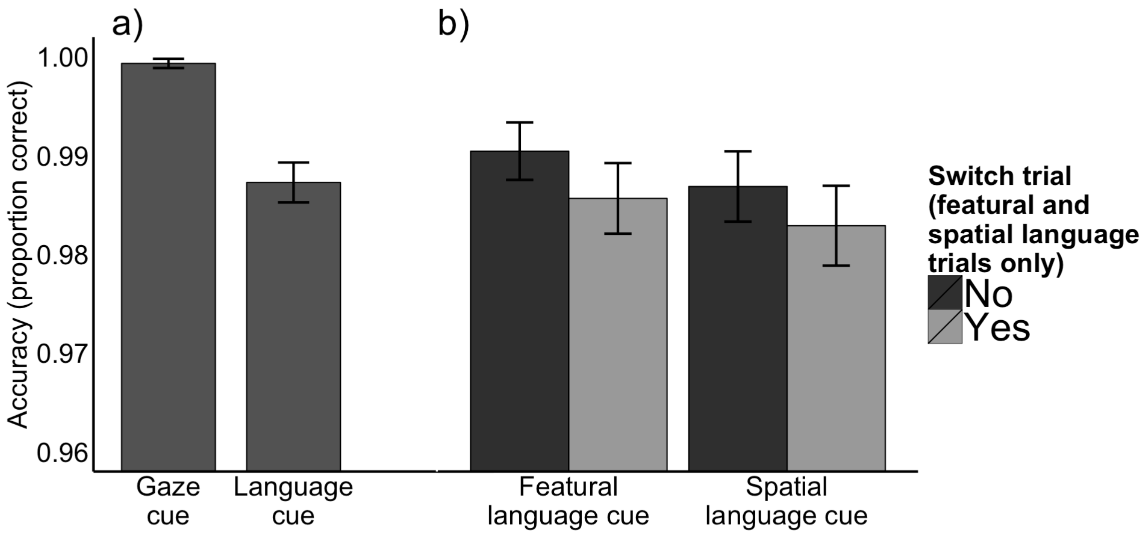

Figure 5.8. Proportion of correct trials a) in the language and gaze cue conditions and b) in the language cue condition, split by the type of language cue and whether or not the trial was a switch trial.

It is possible that participants' accuracy in the language condition was skewed by performance in either the featural or spatial condition. If a reference type effect was present, this may have been affected by whether or not the trial was a switch trial (Figure 5.8b). Specifically, switching from a spatial gaze cue to a spatial language cue may have resulted in different costs than switching from a spatial gaze cue to a featural language cue.

An LMM (Model 5.13), with reference type and presence of a switch as fixed factors and subject and item as random factors, showed no effect of reference type, $\beta = -.574$, $SE = .639$, $z = -.899$, $p = .369$, switch trial, $\beta = -.436$, $SE = .431$, $z = -1.012$, $p = .311$, nor any interaction, $\beta = .303$, $SE = .732$, $z = .413$, $p = .679$. Since there was no evidence that reference type or switching affected accuracy, it is likely that the slightly weaker performance in the language cue condition was indeed due to participants mishearing a small number of language cues, which they were equally likely to do regardless of switch or reference condition.

First saccade direction. The proportion of trials with initial saccades in the correct direction is shown in Figure 5.9a. An LMM (Model 5.13), using type of cue as a fixed factor and object and subject as random factors, showed a significant difference across cue condition, $\beta = -2.572$, $SE = .308$, $z = -8.351$, $p < .001$. This difference, like the accuracy difference, may be due to participants mishearing the language cueing word on some trials. Additionally, it may be due to a higher number of trials in which the participant launched a saccade before interpreting the cue. If language cues did take longer to process than gaze cues, then more of these mistakes would be expected in the language cue condition, resulting in a lower proportion of initial saccades in the correct direction.

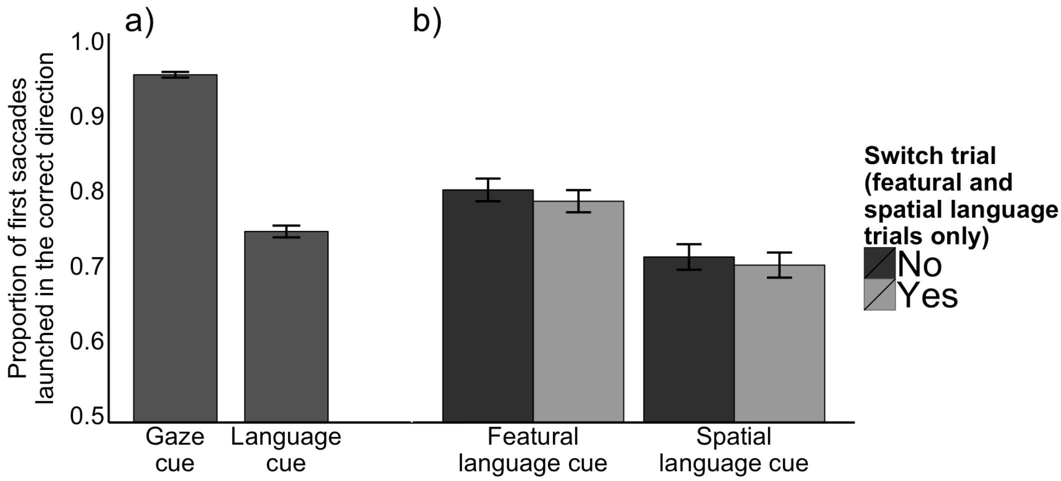

Figure 5.9. Proportion of trials with initial saccades in the correct direction a) in the language and gaze cue conditions and b) in the language cue condition, split by the type of language cue and whether or not the trial was a switch trial.

Figure 5.9b shows the proportion of trials in which the initial saccade was in the correct direction for spatial and featural language cues in switch and non-switch trials. An LMM (Model 5.14), with reference type and presence of a switch as fixed factors and subject and item as random factors, showed a significant effect of reference type, β = -.482, *SE* =.114, z = -4.228, p < .001, but not switch trial, β = -.080, *SE* = .093, z = -.864, p = .387, nor any interaction, β = .038, *SE* = .176, z = .472, p < .637. The increased number of correct first saccades in the featural language condition compared to the spatial language condition may have been due to the increased time taken to process spatial cues. In order to respond to a spatial language cue, the participants were required to process the word ("left" or "right"), then select a frame of reference (in this case, left or right of the central face stimulus from the participants' perspective), whereas in the featural language condition, participants only needed to interpret the word (a colour), then identify the visual stimulus that matched the words meaning, irrespective of any frames of reference. Alternatively, this difference may have been due to more trials in which participants mixed-up left and right, than trials in which participants mixed-up two distinct colours.

First saccade latency. In the language condition, first saccade latency was longer than first saccade latency in the gaze condition (Figure 5.10a). An LMM (Model 5.15), using cue type as a fixed factor and subject and item as random factors, confirmed the difference as significant, $\beta = 132.81$, $SE = 24.43$, $t = 5.437$, $p < .001$. These findings provide more evidence that participants responded faster to the gaze cue stimuli than the language cue stimuli.

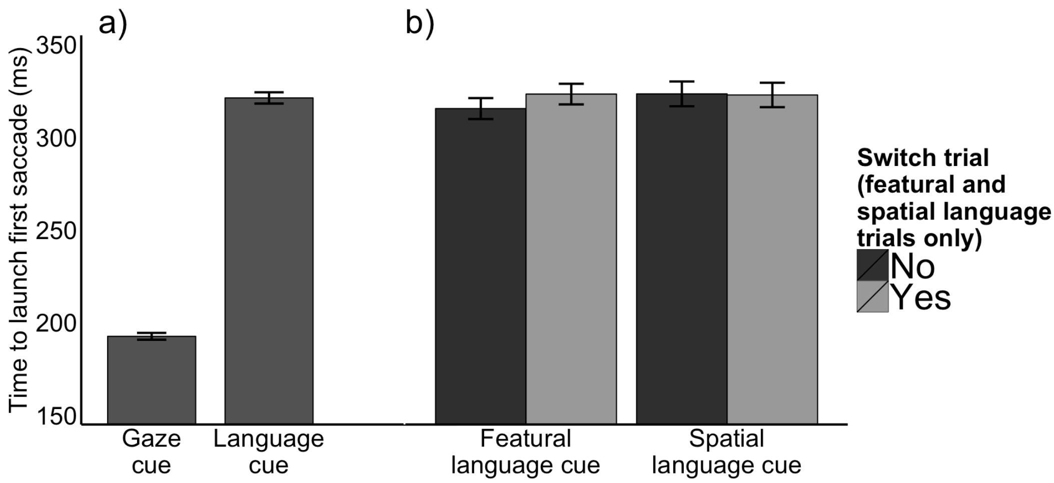

Figure 5.10. Mean first saccade latency in a) the language and gaze cue conditions and b) the language cue condition, split by the type of language cue and whether or not the trial was a switch trial.

An LMM (Model 5.16), with reference type and presence of a switch as fixed factors and subject and item as random factors, showed no effect of reference type, $\beta = 6.301$, $SE = 13.184$, $z = .478$, $p = .619$, switch trial, $\beta = 3.062$, $SE = 7.810$, $z = .392$, $p = .683$, nor any interaction, $\beta = -7.020$, $SE = 15.808$, $z = -.444$, $p = .648$. These results show that neither the type of linguistic reference used nor the presence or absence of a switch in the trial had any significant effect on the time to launch the first saccade.

Time to first fixate the target (FT). Figure 5.11a shows that participants took longer to first fixate the target in the language condition compared to the gaze condition. An LMM (Model 5.17), using cue type as a fixed factor and subject and item as random factors, confirmed the difference as significant, $\beta = 211.75$, $SE = 29.220$, $t = 7.248$, $p < .001$. This result is clear evidence that gaze cues directed participants to the target more quickly than language cues. These findings are crucial for the interpretation of the results from Studies 3 and 4 outlined in the *General Discussion*.

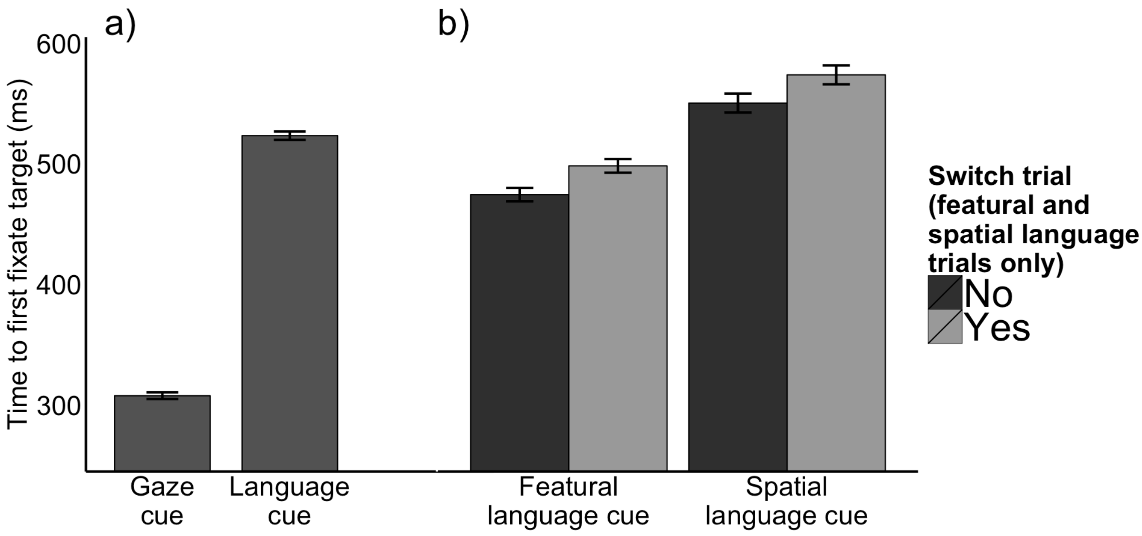

Figure 5.11. Mean time to first fixate the target (FT) in a) the language and gaze cue conditions and b) the language cue condition, split by the type of language cue and whether or not the trial was a switch trial.

It is clear from Figure 5.11b that neither featural nor spatial language cues were responded to as quickly as gaze cues, however responses appear to be slower for spatial language trials compared to featural language trials. An LMM (Model 5.18), using reference type and presence of a switch as fixed factors and subject and item as random factors, found significant effects of reference type, $\beta = 76.011$, $SE = 13.264$, $t = 5.731$, $p < .001$, the presence of a switch, $\beta = 21.853$, $SE = 6.087$, $t = 3.590$, $p < .001$, but no interaction between these factors, $\beta = 14.521$, $SE = 13.397$, $t = 1.084$, $p = .280$.

The results from the LMM confirm the significance of the difference between spatial and featural language trials seen in Figure 5.11b, and also show that participants were slower to fixate the target in switch trials. However, there was no difference in the switching costs between featural and spatial language cue conditions. It appears from these results that participants found switching from a gaze cue to a featural language cue no more difficult than switching to a spatial language cue. However, participants fixated the target more quickly overall following a featural language cue compared to a spatial language cue. As discussed in Chapter 4, in this lab-based paradigm participants may have been able to identify an item by its colour using peripheral vision, meaning that participants did not need to move their eyes to find the target. Spatial language cues did not require visual search either, but these cues did require participants to select a frame of reference from which to move left or right, whereas featural language cues only required participants to match a visual stimulus to a colour word. The frame of reference selection may explain why participants were slower with spatial language cues.

Response time (RT). Like the time to first fixate the target, response time was slower in the language condition compared to the gaze condition (Figure 5.12a). An LMM (Model 5.19), using cue type as a fixed factor and subject and item as random factors, confirmed the difference as significant, $\beta = 187.86$, $SE = 29.260$, $t = 6.420$, $p < .001$. This provides further evidence that gaze cues directed participants to the target more quickly than language cues and that this led to quicker responses. As with the time to first fixate the target results, these findings are important for the interpretation of the results in Studies 3 and 4 because it is clear that the difference in time to process the gaze and language cues influenced the behaviour of participants in these previous studies.

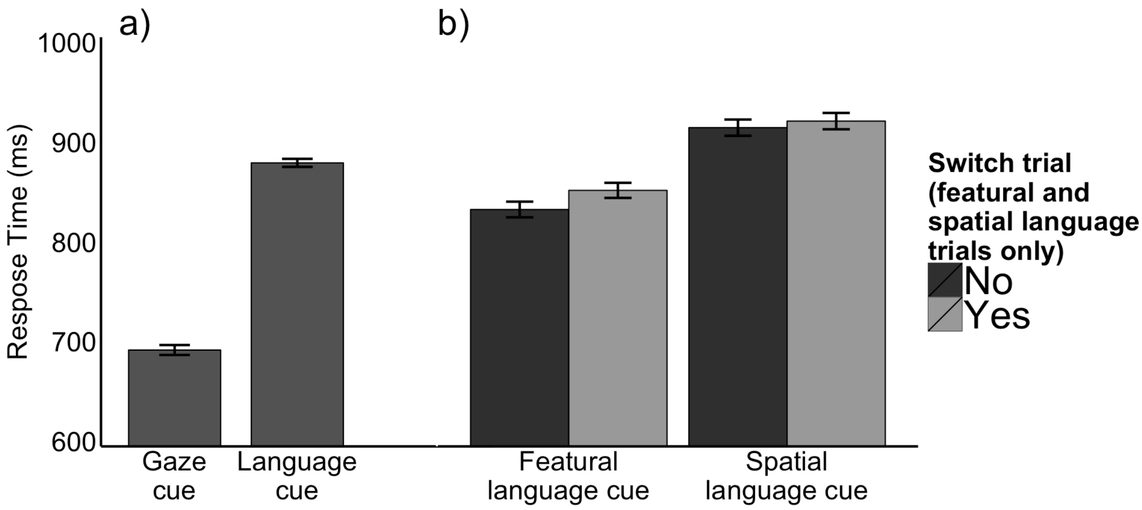

Figure 5.12. Mean response time (RT) in a) the language and gaze cue conditions and b) the language cue condition, split by the type of language cue and whether or not the trial was a switch trial.

Figure 5.12b shows that neither language reference type led to response times as short as the mean response for gaze cues, but like the time to first fixate the target, RT was slower for spatial language trials compared to featural language trials. An LMM (Model 5.20), using reference type and presence of a switch as fixed factors and subject and item as random factors, found significant effects of reference type, β = 77.042, *SE* = 25.939, t = 2.970, p = .006, and an approaching significant effect of the presence of a switch, β = 13.204, *SE* = 6.953, t = 1.899, p = .064. No interaction was found between these factors, β = 11.352, *SE* = 15.143, t = .750, p = .454.

The switch effect found in the time to first fixate the target analysis was only approaching significance for response time. This could possibly be due to the detrimental effect of switching diminishing over time in each trial. Therefore, the effect is less prominent for response time, a later measure than time to first fixate the target. Both the time to first fixate the target and response time results show that participants are slower in language cue trials with spatial references compared to those with featural references. The reason could lie in the particular words used; It may be easier for participants to mix up “left” and “right” than “blue” and “red”. Additionally, even though participants were told that the cues would be from their perspective, there may have been some disruption from participants simulating the perspective of the face on the screen. Although there was no significant effect of reference in experiments 3 and 4 on overall performance, featural trials were slightly slower, so it is interesting that the opposite effect was found here. To speculate, it may be that the centrally fixating face cued visual spatial attention to the centre of the screen, which caused disruption alongside spatial verbal cues, but did not disrupt verbal featural cues.

Overall, Study 5 has shown that participants were quicker to respond to the visual gaze cues used in Study 4 than the language cues used. These findings are key to our interpretation of the results from the previous two studies in this thesis.

General Discussion

Studies 4 and 5 were designed to further investigate the effects found in Study 3. Combined, these three studies explored the effects of varying the reliabilities of concurrent language and gaze in an instructional task. In Study 3, participants were found to favour language cues over gaze cues when reliabilities were equal, despite gaze cues being responded to more quickly. Study 4 supported findings from Study 3, by showing that when gaze cues were subordinate to perfectly accurate language cues, the extent to which the gaze cues influenced responses depended on their reliability. The same effect was not found for subordinate language cues, however the difference in processing time between gaze and language cues may account for this. In Study 3 subordinate language did provide a unique finding: participants were slower when language cues were less reliable, regardless of congruity. This finding was not replicated in Study 4, possibly due to the language cues being used out of the context of a natural instructional sentence. Despite the issues with onset times, these studies showed distinct effects of varying gaze and language cue reliabilities.

The first finding reported from Study 3 was that participants based their responses on the most reliable cue available to them. When gaze and language reliabilities were equal and more reliable than chance, participants showed a preference for following the language cue. This finding is particularly interesting in light of the results of Study 5. Participants were found to respond to isolated gaze cues on average about 200 ms quicker than they responded to isolated spoken language cues, showing that following gaze cues would be an optimal strategy if reliabilities were equal. Furthermore, the results of Study 5 were found using the stimuli from Study 4, where the onset times of the gaze cues and verbal cues were identical, whereas in Study 3 (in which the language preference was found) the onset of the language cue was after the onset of the gaze cue, therefore the potential advantage (in terms of allowing an earlier response) of using a gaze cue over a language cue was even greater than that found in Study 5. It could be argued that Study 3 artificially encouraged a preference for the language cue, as the language used took the form of a direct and explicit instructional command, whereas the gaze cues were subtler. However, this difference between the gaze stimuli and language stimuli is reflective of the use of these cues in natural interactions, with language communicating potentially very detailed and specific information and gaze cues indicating the location the speaker's focus of attention.

When focussing specifically on the sessions in which language was 100% reliable and the gaze cues less reliable, both Studies 3 and 4 found that participants performed slightly, but significantly, worse in trials with incongruent gaze cues when these cues were 80% reliable compared to 50% reliable. These findings suggest that the reliability of the gaze cues change the extent to which these cues are followed, even though language cues are more reliable (and completely reliable). These findings seem at the odds with the results of the real world experiment in Macdonald and Tatler (2013), which found that gaze cues were ignored when language provided enough information for successful task completion. However, a key difference in the present paradigm is that the gaze stimuli were presented centrally, with no need for the participant to seek out the cues. It may be that the highly flexible effects found in Macdonald and Tatler (2013) were predominantly driven by the gaze seeking stage of gaze utilisation, which was not present in lab-based Studies 3, 4 and 5. Alternatively, it is possible that the larger gaze stimuli (in terms of the proportion of the visual field that the gaze cue stimuli occupy) in studies 3, 4 and 5 may have been harder for the participants to actively ignore.

The effects of incongruent gaze cues in these sessions were not, of course, entirely automatic, as the reliability affected the extent to which the gaze cues were followed. Given the preference shown for language in Study 3 when gaze and language cues were equally reliable, it is surprising that an incongruent gaze cue would affect accuracy at all when language was more reliable. However, it should be noted, when cues were incongruent to each other, participants selected the target cued by 100% reliable language over 90% of the time when gaze was 80% reliable in Studies 3 and

4. This is higher than when the cues were both 80% reliable in Study 3 (around 75%). The additional detrimental effect of incongruent gaze cues when they were more reliable is in line with findings from gaze-cueing paradigm studies that have manipulated cue reliability. Hill et al (2010) and Kuhn and Kingstone (2009) found that after 500 ms and 900 ms respectively, participants would be influenced by the perceived reliability of a gaze cue. In the current paradigm, the response times show that there was typically enough time for top down influence of perceived reliability to affect the utilisation of the gaze cues and consequently accuracy. The difference in time to process and respond to the different cues may also have contributed to the increased detrimental effect of 80% reliable incongruent gaze cues; the gaze cues were interpreted earlier than the language cues (as shown in Study 5), so participants may have been more likely to respond to the gaze cue before processing the language cue when the gaze cues were 80% reliable compared to when they were 50% reliable.

The results of our first saccade direction measure, the time to first fixate the target and response time were all very similar in Studies 3 and 4 for the sessions with 100% reliable language and less reliable gaze cues. Both studies found that when gaze cues were more reliable they benefitted performance when congruent and hindered performance when incongruent, again providing evidence that the gaze cues were utilised more when they were more reliable and showing that the findings in Study 3 were not driven by a misalignment of cue onset times.

The first saccade latency and the time to first fixate the target differed between Studies 3 and 4 for sessions with 100% gaze reliability and less reliable language cues, but not in the way that was anticipated. The lack of an effect of incongruity between the cues in Study 3 was interpreted as being due to the later onset of the language cue resulting in the incongruent language cues not having enough time to negatively affect performance. It was hypothesised that in Study 4, with the onset times of the two cues synchronised, a congruity effect would be found. This, however, was not the case. There was no effect of congruity, suggesting that participants were still able to process and respond to the gaze cues more quickly than the language cues, which was confirmed with the results of Study 5. The difference found between Studies 3 and 4 was that in Study 3 there was an overall detrimental effect of 50% reliable language compared to 80% reliable language, whereas no such effect was found in Study 4. It is possible that with the low number of participants (five) in Study 3, the detrimental effect was peculiar to these particular participants and that the more reliable results of Study 4 (with ten participants) show the true effect. Alternatively, it may be that the difference in the context of the critical cueing word caused the differing results. In Study 3 the critical (disambiguating) word was presented in a full natural instructional sentence, whereas in Study 4 the critical word was presented in isolation after the word “Select”. It is possible that hearing incorrect full instructional sentences repeatedly, but not incorrect words in isolation, slows down performance.

The response times also differed between Studies 3 and 4 for sessions with 100% gaze reliability and less reliable language cues. In Study 4 there was a significant effect of congruity on response time, with incongruent language cues slowing down responses even though the gaze cues were always correct. This finding supported the hypothesis that a congruity effect would be found when the cue onset times were synchronised. However, the lack of a congruity effect for language in the time to first fixate the target suggests that gaze cues were still responded to more quickly, as saccades to the gazed-at locations appear to have been initiated before incongruent language cues were interpreted. This difference in processing time was confirmed by the results of Study 5. The congruity effect shows that like gaze cues, language cues can be distracting even if another cue is more reliable. In Study 4, there is no evidence that the distracting effect of incongruent language cues was mediated by their reliability, but this possibility cannot be discounted due to the extra time required to process and respond to language cues compared to gaze cues.

The extra time taken to respond to language cues was likely due to the additional processing required with a language cue compared to a visual spatial cue like gaze. Firstly, linguistic cues must initially be processed and interpreted. Secondly, there is the issue of perspective. Gibson and Kingstone (2006) carried out a series of experiments comparing a number of visual cues, including gaze cues and written spatial language cues. Although the language cues were visual rather than auditory, the comparisons made by Gibson and Kingstone remain relevant to the present findings. The authors proposed a new way of categorising visual spatial cues: "projective" cues and "deictic" cues. Projective cues, such as the words "left" or "right" require participants to select a frame of reference for the cue in order to determine the direction being cued. Gibson and Kingstone argue that deictic cues, such as arrows and gaze, specify a target directly with regard to the viewer, with no need for the participant to select a reference frame. The auditory spatial language cues in the current paradigm are analogous to the projective spatial cues discussed by Gibson and Kingstone. It is therefore possible that the lexical processing of the words "left" or "right", combined with the selection of a reference frame made responses to language slower than responses to gaze.

The distinction between projective and deictic cues, although proposed with regard to visual spatial cues, may also provide insight into why participants were quicker to respond to featural language cues in isolation compared to spatial language cues. Like the spatial language cues, featural language cues required lexical processing in order to identify the colour of the target item. However, unlike spatial language cues, there was no need for participants to select a frame of reference to find the target; a cue to a target of a particular colour would always be interpreted as the same target, regardless of the perspective taken. The absence of the perspective selection stage could account for the quicker responses to featural language cues compared to spatial language cues. This may initially seem at odds with the design of Study 2, in which featural references were considered “harder” than spatial references (in that gaze cues were of more use in the featural reference trials than the spatial reference trials). However, as discussed in Chapter 4, in the present lab-based paradigm, the target could have potentially been identified using peripheral vision after a featural cue as there were only two targets taking up a large proportion of the participants’ visual field. In Study 2, there were 60 items, several of which shared colours, so participants had to visually search in order to identify the correct item. The featural language cues therefore played different roles in the naturalistic Study 2 compared to the lab-based Studies 3, 4 and 5. In the former, the featural language cues supplied reliable information distinct from spatial gaze cues, thus making these references more useful than the spatial language cues that supplied overlapping information with the gaze cues. In the latter studies, featural cues provided information that did not require the participants to establish external frames of reference, unlike the spatial language cues.

The two studies presented in this chapter used controlled, simple gaze and language stimuli to extend the investigation into the effects of varying the reliability of concurrent gaze and language cues. The real world studies used in this thesis have found an interaction between gaze and language; with gaze being used more when it provides more uniquely useful information. In these lab-based studies, the reliabilities of language and gaze have been varied in a controlled and systematic way to provide a more detailed picture of how language and gaze cue reliabilities affect each other. In terms of accuracy, participants strategically used the most informative cue available to them. When cues were equally informative participants showed a bias for language, despite the extra time required to process these cues. Gaze cue effects were, however, still persistent. When gaze cues were subordinate to completely reliable language cues they caused distraction, and the degree of this distraction was mediated by the gaze cue reliability. Language cues appear to take longer to process, limiting the interpretation of their distracting effects in this paradigm. However, for language cues embedded in a natural instruction sentence, an overall distracting effect was found from unreliable language when gaze was 100% reliable regardless of the cue congruity in any particular trial. No such effect was found with gaze cues. Combined, these studies show some clear differences between these two cues. Gaze cues affect attention relatively quickly, but the extent of their attentional effects can also be quickly mediated by their perceived reliability, whereas language cues appear to be preferred, despite taking longer to process. Furthermore, low language reliability was found to disrupt overall performance, but only when the cues were embedded in a natural instructional sentence.

Chapter 6:

Discussion

The five studies that make up this thesis investigated the relationship between gaze cues and language in natural communication. In the present chapter, the results of all of these studies are summarised. The effects of language on gaze seeking and gaze following are discussed in relation to these experimental results, focusing specifically on the novel theoretical contributions of these findings. It is argued that the utilisation of gaze cues in natural communication is largely influenced by concurrent spoken language.

Summary of experimental results

In Study 1 the presence of social roles was manipulated in a collaborative task while eye movements were measured. In pairs, participants worked together to make a cake. Half of the pairs were given roles ("Chef" or "Gatherer") and the other half were not. Contrary to the results of static image experiments, participants spent very little time looking at each other, challenging the generalisability of the conclusions from lab-based paradigms and extending the previous findings that social presence affects the extent to which we look at others (Laidlaw et al., 2011; Gallup, Hale et al., 2012) to real interactions, showing that this behaviour is not simply a way to prevent an interaction occurring with a stranger. Participants were found to instead spend much more time looking at the same items (shared gaze). When given spoken instructions, listeners in the roles condition looked at the speaker significantly more than listeners in the no roles condition. These findings suggest that our tendency to seek the gaze cues of collaborators is affected either by our social perceptions or our perception of the reliability of the collaborator. Furthermore, the analysis of these interactions found participants used a mixture of spatial and non-spatial verbal references, which influenced gaze behaviour.

Study 2 used a controlled real-world paradigm to investigate the effect of the type of verbal reference used in an instruction on gaze behaviour. Specifically, whether the use of featural or spatial disambiguating determiner words affected the extent to which participants used inherently spatial gaze cues. Each participant followed instructions to complete a real-world search task, while wearing a portable eye-tracker. The instructor varied the determiner used (featural or spatial) and the presence of gaze cues (absent, congruent or incongruent). Fixations to the objects and instructor were recorded. Congruent gaze cues were used and followed more when they were provided alongside featural references. Incongruent gaze cues were not initially followed any more than chance. However, unlike participants in the no-gaze condition, participants in the incongruent condition did not benefit from the use of spatial instructions over featural instructions. These findings suggest that although participants selectively used informative gaze cues and inhibited gaze following when cues were unreliable, visual search was nevertheless disrupted when inherently spatial gaze cues were accompanied by contradictory verbal spatial instructions.

The final three studies in this thesis used a lab-based paradigm to further investigate the effects of varying the reliabilities of concurrent language and gaze cues. The reliabilities of language and gaze were varied in a controlled way to provide a more detailed picture of how language and gaze cue reliabilities affected each other. Participants were found to strategically use the most reliable cue available to them and when these cues were equally reliable participants showed a bias for language, despite the extra time required to process these cues, suggesting the language cues were the dominant cue. More evidence for the dominance of language was provided by results showing that when language cues were embedded in a natural instruction sentence, an overall distracting effect was found from unreliable language when gaze was 100% reliable regardless of the cue congruity in any particular trial. No such effect was found with gaze cues. Gaze cue effects were, however, still present and persistent at low reliabilities. When gaze cues were subordinate to completely reliable language cues they were still initially followed. However, the extent to which subordinate gaze cues were followed was mediated by their reliability; when these cues were 80% reliable they were followed more than when they were 50% reliable. Combined, these three studies show some clear differences between these two cues. Gaze cues affected attention relatively quickly, but the extent of their attentional effects were also quickly mediated by their perceived reliability, whereas language cues appeared to be preferred, despite taking longer to process. Furthermore, low language reliability was found to disrupt overall performance, but only when the cues were embedded in a natural instructional sentence.

The interaction between spoken language and gaze cues

The combined results of all five studies in this thesis contribute to our overall understanding of how gaze cues and language interact in natural communication. To discuss these contributions, the present section outlines the two key stages involved in a gaze cueing event chronologically and explores the factors that influence behaviour in each of these stages, with specific focus on the novel theoretical contributions of findings in this thesis.

Looking at the eyes of another

As discussed in depth in Chapter One, there has long been eye-tracking evidence showing that people have a tendency to look at the eyes of another in paintings (Buswell, 1935) and photographs (Yarbus, 1967). This preference has been shown to persist when accounting for central fixation biases (Bindemann et al., 2009) and biases for fixating the centre of a face (Levy et al., 2013), suggesting that the tendency to fixate on the eyes of another is not simply a result of face recognition processes. Studies using more complex static stimuli, with multiple people in a single image, have shown that participants are more likely to fixate the eyes of people than non-biological social cues, such as arrows (Birmingham et al., 2009) and that participants are more likely to look at eyes when the people in the image are interacting with one another (Birmingham et al., 2007). The results of the former study provides evidence that eye gaze is a more attractive social stimulus than arrows, and the results of the latter study suggest that eyes are looked at more when they potentially have communicative value. All of these experiments however, lacked social presence; the gaze cues were provided by still images of people, rather than a present human being. In the real world, gaze-orienting behaviour is markedly different than it is in still image studies, limiting the usefulness of static image paradigms in this research area. Study 1 in this thesis showed this difference clearly; the overall time participants spent looking at each other was between 2-4% of task time.

Laidlaw et al (2011) compared looks to another between a present person in a waiting room and a person on a monitor. Participants were found to look significantly more at the person on the monitor, suggesting that social presence decreases the extent to which we look at another. Furthermore, there was a positive correlation between the social skills of the participants (as assessed with the Baron-Cohen et al. (2001) autism quotient questionnaire) and the extent to which participants looked at the present confederate, suggesting that those with better social skills may be more comfortable looking at another. It is important to note however, that in this situation the other person was a stranger, whereas in natural communication, which is present in Study 1, the other person is necessarily involved in a social interaction with the viewer. Therefore, although the influence of social effects in the waiting-room paradigm informs us about social influences on how we look at others, the results are not directly comparable to the findings of the present thesis, which explored the interaction between gaze and language in communicative interactions.

This thesis is certainly not the first to consider natural gaze behaviour during interactions, as a great deal of work was published by social psychologists throughout the nineteen-sixties and –seventies. Argyle and Dean (1965) found that the extent to which we are comfortable looking at a person is influenced by our proximity to that person, with people more resistant to eye contact at close distances. These results demonstrate a clear social effect on how we look at others during interactions. The aforementioned small proportion of time spend on partner gaze in Study 1 suggests that similar social effects were affecting behaviour while people baked together. The participants, who had mostly never met before, were perhaps following social norms and avoiding looking directly at each other. Despite the pairs spending very little time looking at each other, there was still evidence of partner gaze when using spoken language, which was affected by the roles of the individuals. During instructions, listeners in the roles condition were more likely to look at the speaker than listeners in the no roles condition or the speakers in both roles and no roles conditions. This suggests that the extent to which we look at a speaker during communication is influenced by our social perceptions of that person; if they are perceived to be of higher social status, we will look more. Although not a direct measure of gaze seeking, the finding that social perceptions affect the extent to which we look at another during an interaction has clear implications for our ability to use and following gaze cues. The more we look at another, the more likely we are to identify their non-verbal cues, including gaze cues.

Study 2 showed that the tendency to avoid looking directly at someone with whom one is interacting could vary from situation to situation. In Study 1 participants were moving about a kitchen working on several different worktops and collecting items from around the whole room. In Study 2 participants stood facing an instructor at the onset of each instruction. In this more controlled real-world study, participants almost always began a trial by looking at the instructor. The only exception was when participants anticipated an unhelpful gaze cue, but even then in over 70% of trials participants were looking at the instructor at the onset of the trial. Due to the participant and instructor facing each other, if the participant wished to look away from the instructor they were required to actively move their eyes. This may account for the high proportion of looks to the instructor in Study 2. Participants appear to have only actively looked away from the instructor when they anticipated that an unhelpful cue was about to be given. Later in the sentence, at the point of the gaze cue, there was more evidence of selective gaze seeking, with participants looking more often at a cue they anticipated to be helpful and less often at the cue they anticipated to be unhelpful. These findings support lab-based studies that have found that people look more at gaze cues they anticipate to be useful (Itier & Villate, 2007) and extend these findings to real-world tasks.

Like some previous research, Study 2 showed evidence of spoken language affecting the extent to which people orient towards gaze cues. Macdonald and Tatler (2013) used a paradigm in which gaze cues were sometimes required to complete the task, as language was ambiguous. In these cases gaze cues were sought out often, but they were rarely sought out when language provided the necessary information to complete the task. Study 2 extends our understanding of the reflexive relationship between gaze cue utilisation and language by showing that the type of language used, rather than the specificity of that language, also affects the use of gaze cues. Participants were found to seek out gaze cues more when these cues were provided alongside featural language cues rather than spatial language cues. Gaze cues are inherently spatial rather than featural; gaze can direct an observer to a location in space but it cannot direct an observer to a colour. Therefore, there was no overlap between the information provided by the spatial gaze cue and the featural language cue, whereas there was an overlap with the spatial language cue. Participants were found to selectively seek out gaze more when gaze cues provided more exclusively useful information. There was no difference in gaze seeking behaviour found between the no gaze and incongruent gaze condition, suggesting that participants simply ignored cues they anticipated to be unhelpful. This finding is particularly interesting given the wealth of evidence showing distracting effects of gaze cues when they cue locations incongruent with a target (e.g., Ricciardelli et al., 2002; Ristic & Kingstone, 2005; Galfano et al., 2012). It is possible that these well-established reflexive effects are unique to the gaze following stage of gaze utilisation, as this is the stage most previous experiments have explored. From the findings of the present thesis, gaze seeking appears to be a largely flexible process.

Study 2 also found evidence that the social skills of participants affected gaze-orienting behaviour and interacted with the type of linguistic reference used. More specifically, the effect of reference type on the time participants spent looking at the instructor was different for the two Autism Quotient (AQ) groups. The high-AQ group did not significantly differ in the extent to which they looked at the instructor between spatial and featural reference trials, whereas the low-AQ group spent significantly longer looking at the instructor when featural references were used compared to spatial references. In Chapter 3 it was speculated that this might have been due to the low AQ group being more sensitive to social norms that discourage looking directly at another when not necessary, thus there was a difference between when the cues were more and less relatively useful. The high AQ group, however, being less sensitive to these social norms, sought out the gaze cue regardless of how useful they were relative to language. Although this is speculative, it is in line with an account provided for results of a previous study showing that low AQ individuals would look at a face of a speaking person on a monitor and the face of a present individual to the same extent, whereas higher AQ individuals were more likely to look at a present person than a person on a monitor (Freeth et al., 2013).

The studies outlined above illustrate that our social perceptions and the language that we use influence the extent to which we look at another and when we look at that other during an interaction. The studies in the present thesis have added to our understanding of the interplay between gaze and language during gaze seeking by showing that people are more likely to look at a speaker when they are of higher social status. This is not only an important finding in itself, but also shows the importance of investigating social attention in real social interactions, as standard gaze cueing paradigm experiments would not be able to identify this effect. Furthermore, these studies have shown that gaze seeking is a highly flexible process, with participants using inherently spatial gaze cues more when the information is more helpful in the task relative to accompanying language.

Following the gaze of another

The gaze following stage is the most well researched stage of gaze utilisation. As, outlined in Chapter 1, much of this research has been conducted using variants of Friesen and Kingstone's (1998) Posner-type (1980) gaze-cueing paradigm. With this highly controlled simple paradigm a great deal of data has been collected that can help us to understand the attentional processes that occur upon seeing a gaze cue. While there have been a number of varied and potentially contradictory findings and conclusions from these numerous studies, there are a few fundamental findings consistent between most gaze cueing paradigm experiments. The studies in the present thesis offer insights into the effects of language on these gaze following behaviours.

The principle finding of Friesen and Kingstone's study (1998) was that unreliable gaze cues benefited performance in a target selection task when the cue was congruent to the target location, even though participants understood these cues to be non-predictive. This finding has been replicated numerous times (e.g., Driver et al, 1999; Langton & Bruce, 1999) and has been extended to show a reciprocal detrimental incongruent gaze cueing effect (e.g., Ricciardelli et al., 2002; Kuhn & Benson, 2007). Eye tracking studies have also shown eye movements to be affected by incongruent gaze cues, in the form of increased saccade errors (Ricciardelli et al., 2002; Galfano et al, 2012). These findings together show that viewing a gaze cue leads to a reflexive shift in attention. In Studies 3 and 4, in which a novel version of the gaze cueing paradigm was used, participants were distracted by subordinate incongruent gaze cues when language was 100% reliable, supporting the view that gaze cues cause an automatic shift in a viewer’s attention. As has been noted previously in this thesis, findings from these simple paradigms may not extend to real world interactions. However, Study 2 also found evidence for an automatic gaze cueing effect in a complex real world environment; participants were found to identify targets faster when provided with a spatial language cue compared to a featural language cue when no gaze cues were provided, however when a gaze cue that participants’ knew to be cueing the incorrect location was provided alongside spoken language, there was no longer a benefit to using spatial language cues. This finding suggests that gaze cues disrupted participants’ processing of spatial language cues, even though participants knew the gaze cues were incorrect.

The issues raised by, amongst others, Ricciardelli et al. (2002) and Tipples (2002) regarding the uniqueness of gaze cue stimuli in Posner-type (1980) tasks are arguably also relevant for the real-world and video tasks used in this thesis. As with the stimuli in the typical gaze-cueing paradigm (Friesen & Kingstone, 1998), some of the seemingly reflexive early responses of the gaze and head cues used in this thesis could arguably have been caused by the visual motion of the head, rather than any unique social relevance of the eyes. In addition, the conventional gaze cueing paradigms explicitly employ gaze cues only, without any potentially confounding head movements. Gaze cues were combined with head movements in this thesis primarily for the sake of ecological validity, however this decision can also be supported by the theoretical insights and findings of Lozincz, Baker and Perrett (1999). Lozincz et al. found that rhesus monkeys will follow the gaze cue of other monkeys when the cues are incongruent to the head, but that they would follow the head direction when incongruent to body posture if eye direction were not clearly visible. From these findings they theorised that there is a hierarchy of cue following, with eye-gaze overriding head direction and head-direction overriding body-position. While it is not clear from the results of Studies 2-5 whether participants explicitly followed eye-gaze-direction, head direction or both, the work of Lozincz et al. (1999) would suggest that this distinction may not be important for the questions being asked in this thesis; head direction and gaze direction both cued the location of the looker's focus of visual attention.

Although there is the above evidence for reflexive gaze following, it is clear from previous literature as well as many of the findings in the present thesis that other factors also influence the extent to which we follow gaze cues during communication. One of these factors is the perceived reliability of a gaze cue. A number of studies have shown that when decreasing the reliability of gaze cue stimuli from 50% (thus making it more likely that a cue will be incorrect than correct) the typical gaze cueing effect persists, but only at small SOAs (Kuhn and Kingstone, 2009; Hill et al., 2010). At longer SOAs, however, these effects were not found, showing that knowledge of the reliability of a cue can override the short-lived attentional shift.

Given that gaze following responses have been shown to be time sensitive, it is worth considering the time course of the gaze following measures used in this thesis for the various paradigms. In Study 2, participants were free to search for the target before and after the gaze cue was given, definitely providing the opportunity for low-level reflexive response to occur. This opportunity was also available in Studies 3, 4 and 5, as participants were free to select a target as soon as the gaze cue was initiated. More important for the present thesis are the cue onset times in relation to the onset of disambiguating phrases. In Studies 2 and 3, gaze cues were provided (on average) 96 ms and 172 ms before the disambiguating language phrase respectively and in Study 4, language and gaze cue onsets were concurrent. Whereas, there was little comparability between gaze seeking measures in Study 2 and Studies 3-5, in terms of gaze following, all of the studies in this thesis using concurrent language and gaze presented language cues within 200 ms of gaze cue onset, allowing for potential reflexive shifts in attention at the time of hearing the disambiguating words (Kuhn & Kingstone, 2009; Hill et al., 2010).

As well as changing the absolute reliability of a gaze cue to investigate gaze following, it is also possible to vary the communicative information provided around the cue. In real interactions the primary communicative tool accompanying a gaze cue is typically spoken language and many researchers have investigated the interaction between language and gaze. Knoeferle & Kreysa (2012) found evidence that the level of difficulty of processing language affected how gaze cues were used; supportive gaze cues were used when concurrent language used a common grammatical structure (in German), but not when language used an uncommon (but grammatically legal) structure. This appears to show that gaze following was influenced by the difficulty in processing language; when language was easy to process, cognitive resources were free to exploit supportive gazes cues, however when language was more difficult to process, these cognitive resources were not available to make use of supportive cues. Macdonald and Tatler (2013) later found that when cues were supportive in a task, but not essential for task completion, they were rarely followed. Gaze cues were followed significantly more when gaze provided essential information, again showing evidence for selective gaze following, based on the informativeness of the gaze cue relative to language. Study 2 of the present thesis provided evidence that gaze following is not only affected by the relative informativeness of language but also the type of language used; when language provided spatial information (which overlapped with the spatial information provided by gaze) participants followed gaze less than when language provided featural information (which did not overlap with information provided by gaze). This is arguably the same effect that was found by Macdonald & Tatler (2013), with gaze cues being followed more when they were more useful relative to language, however there was a key difference between these studies: in Study 2 all sentences were unambiguous. In Macdonald & Tatler (2013) it is possible that the increased gaze following when language was ambiguous was due to participants seeking out supporting information simply because of confusion at the ambiguous instructions, however the present thesis has shown that this flexible behaviour was actually related to the information gaze cues provided, rather than the lack of information provided by language.

Many of the findings in this thesis show that the context around a gaze cue, and more specifically the language around a gaze cue affects how these cues are utilised. While these contributions are novel, other researchers have previously theorised on the importance of communicative intent on gaze following. Senju & Csibra (2008) found that infants were more likely to follow the gaze cue of an adult if the cue was preceded by another communicative cue, such as direct eye contact or speech. Senju & Csibra (2008) theorised that the early development of gaze following behaviour is supported and perhaps even reliant on other communicative tools. The results of the present thesis show the language used around a gaze cue does more than shape the development of a gaze following mechanism in infancy, but is rather key to the way adults utilise the gaze of another in a natural interaction.

The findings outlined above suggest that although gaze cues lead to immediate shifts in attention, in natural interactions gaze following is heavily influenced by the content of concurrent language. The extent to which we follow gaze cues appears to be influenced by the form and content of our primary communicative tool: spoken language. However, it is important to consider that the paradigms that report these findings have perhaps set gaze cues up to be supportive cues. In Knoeferle and Kreysa (2012), Macdonald and Tatler (2013) and Study 2 of this thesis participants were not told of the presence of gaze cues prior to the experiment, but they were informed that they would hear spoken sentences, and in the case of Macdonald and Tatler (2013) and Study 2 participants were explicitly told to follow spoken instructions. In Study 3, a paradigm was used to make language and gaze cues equivalent, except in terms of reliability, which was manipulated in a controlled manner. When language and gaze cues we equally reliable, more reliable than chance and incongruent to each other, participants selected the target cued by language in most trials even though following the gaze cue would have been quicker. This provides support to the idea that language cues are favoured over gaze cues, which provide a supportive role to language.

It is clear that when viewing a gaze cue there is a tendency to attend to the gazed-at-location. While evidence that this effect is unique to gaze cueing stimuli (Ricciardelli et al, 2002, Ristic & Kingstone, 2005) may be outweighed by evidence that this effect is shared by a number of visual spatial cues (Tipple, 2002; Galfano et al., 2012), there are a number of studies showing that the gaze cueing effect is stronger than the similar effect of other visual-spatial cues (Quadflieg, Mason & Macrae, 2004; Downing et al., 2004; Kuhn & Benson, 2007). Regardless of whether this effect is unique to gaze or not, Studies 2, 3 and 4 of this thesis have found evidence to support the presence of the automatic gaze cueing effect. However, the present thesis has also found evidence showing the flexible nature of gaze following, and specifically the influence of language. Participants were found to follow gaze cues more when these cues were more informative relative to concurrent language, showing a top down strategic use of gaze cues. Taken together, the gaze following results from the present thesis show that although the immediate attentional effects of a gaze cue are present in real interactions, our responses to gaze cues are largely driven by our perception of the reliability of these cues, which can be modulated by concurrent spoken language.

Future directions for gaze cueing and language research

The findings of this present thesis raise a number of interesting questions about the flexible interaction between language and gaze cues. As a direct follow-up to the latter three studies of this thesis, an experiment could be carried out using the basic framework of Studies 3 and 4, but offsetting the cue onsets in order to align processing completion times. To do this, an experiment using one cue at a time (as in Study 5) could be carried out and the mean time difference between gaze and language responses could be calculated. This mean time difference could be used to offset the onset times of gaze and language cues to increase the likelihood that both cues are interpreted at the same time. This paradigm would allow insight into the extent to which the distinct effects of subordinate language and gaze cues in Studies 3 and 4 were due to the processing times of the respective cues.

Another line of future research would be to explore whether the relationship between gaze and spoken language is unique to these two factors, or whether gaze cue utilization is similarly flexible with other concurrent cues. Likewise, are there cues other than gaze cues that are used in the same way to support ambiguous language? It may be that language is a prevalent naturally occurring modulator of gaze cue reliability/informativeness, or it may be that there is a unique relationship between these two communicative cues that has developed to aid communication. There are other non-verbal cues that have been investigated within the fields of social attention and natural communication, such as pointing (Daum, Ulber, & Gredebäck, 2013) or hand gestures (Yu & Smith, 2013). There is scope for comparative research between gaze cues and these other social cues to investigate whether the communicative gaze of another really does have a unique relationship with spoken language.

The effects of social roles on gaze orienting found in Study 1 lead to interesting questions about the underlying cause: Did listeners look more at speakers who were Chefs because of the perceived social status of their partner or their perceived informativeness? Future experiments could dissociate between the social status and informativeness of a collaborator. For example, in a 2×2 design, participants could collaborate on a task with a partner (a confederate) who would either be introduced as high or low social status (e.g., a lecturer or fellow student) and the participant would be informed of the confederate's expertise in the task (e.g., a novice or an expert). Dependent variables would be eye movement measures including number of looks to the confederate (and total looking time).

While the highly naturalistic Study 1 provided data relating to the extent to which we look at others during real interactions, it did not directly measure how people follow gaze cues. Future studies could investigate how our social perceptions of another affect not only how we look to that person, but the extent to which we follow their cues and remember where they look. This could be achieved by using paradigms similar to the real world task in Study 2, but with the inclusion of manipulations of social perceptions.

Summary

The present thesis explored the relationship between language and gaze cues in communication. Much of the previous literature on gaze cue utilisation used artificial stimuli that arguably reduced gaze cues to simple visual spatial cues that lacked key characteristics of natural gaze. For this reason, the present thesis began with a largely observational study, allowing for spontaneous natural language and gaze. Participants were found to rarely look at their partners, but to do so strategically, with listeners looking more at speakers when the latter were of higher social status. Eye movement behaviour also varied with the type of language used in instructions, so Study 2 used a more controlled, but still real-world paradigm to investigate the effect of language type on gaze utilisation. Participants used gaze cues flexibly, by seeking and following gaze cues more when they were accompanied by distinct featural verbal information compared to overlapping spatial verbal information, showing that gaze cue utilisation is related to the information gaze cues provided, rather than the lack of information provided by language. The final studies reduced gaze and language cues to equivalent stimuli and varied the reliability of each. Even in this artificial paradigm, language was preferred when cues were equally reliable, supporting the idea that gaze cues are supportive to language. Typical gaze cueing effects were still

found, however the size of these effects was modulated by gaze cue reliability. Combined, the studies in this thesis show that although gaze cues may automatically and quickly affect attention, their use in natural communication is mediated by the form and content of concurrent spoken language.

References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, **73**, 247-264.

Anderson, N. C., Risko, E. F., & Kingstone, A. (2011). Exploiting human sensitivity to gaze for tracking the eyes. *Behavioural Research*, **43** (3), 843-52.

Argyle, M. (1988). *Bodily Communication* (Second Edition). London: Methuen.

Argyle, M. & Cook, M. (1976). *Gaze and Mutual Gaze*. Cambridge: Cambridge University Press.

Argyle, M., & Dean, J. (1965). Eye-contact, distance and affiliation. *Sociometry* **28**, 289-304.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: MIT Press.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-spectrum Quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, **31** (1), 5-17.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random-effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language,* **68***,* 255-278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-5. [Software]. Available from: http://CRAN.R-project.org/package=lme4

Bayliss, A. P., di Pellegrino, G. & Tipper, S. P. (2005). Sex differences in eye gaze and symbolic cueing of attention. *Quarterly Journal of Experimental Psychology*, **58A** (4), 631-650.

Bayliss, A. P., Griffiths, D. & Tipper, S. P. (2009). Predictive gaze cues affect face evaluations: The effect of facial emotion. *European Journal of Cognitive Psychology*, **21** (7), 1072-1084.

Bayliss, A. P. & Tipper, S. P. (2006). Gaze cues evoke both spatial and object-centred shifts of attention. *Perception and Psychophysics*, **68** (2), 310-318.

Bindemann, M., Scheepers, C., Burton, A. M. (2009). Viewpoint and center of gravity affect eye movements to human faces. *Journal of Vision*, **9** (2):7, 1-16.

Birmingham, E., Bischof, W., & Kingstone, A. (2007). Why do we look at people's eyes? *Journal of Eye Movement Research*, **1** (1):1, 1-6.

Birmingham, E., Bischof, W., & Kingstone, A. (2008). Social attention and real-world scenes: The roles of actione, competition and social content. *The Quarterly Journal of Experimental Psychology*, **61** (7), 986-998.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2009). Get Real! Resolving the debate about equivalent social stimuli. *Visual Cognition*, **17** (6), 904-924.

Birmingham, E., & Kingstone, A. (2009). Human social attention: A new look at past, present and future investigations. In A. Kingstone and M. B. Miller (Eds), *Annals of the New York Academy of Sciences: The Year in Cognitive Neuroscience 2009*, New York: Wiley.

Brennan, S. E., Hanna, J. E., Zelinsky, G. J., & Savietta, K. J. (2012). Eye gaze cues for coordination in collaborative tasks. *DUET 2012 Workshop: Dual eye tracking in CSCE. 2012 ACM Conference on Computer Supported Cooperative Work*, Seattle, WA.

Brown-Schmidt S., & Tanenhaus M. K. (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science*, **32** (4), 643–684.

Buswell, G. T. (1935). *How People Look at Pictures.* Chicago, IL: University of Chicago Press.

Clark, H.H. and Krych, M.A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, **50** (1), 62-81.

Cooper, R. M. (1974). Control of eye fixation by meaning of spoken language: New methodology for real-time investigation of speech perception, memory and language processing. *Cognitive Psychology*, **6**, 84-107.

Crosby J. R., Monin B., & Richardson D. R. (2008). Where do we look during potentially offensive behaviour? *Psychological Science*, **19** (3), 226-228.

Daum, M. M., Ulber, J., & Gredebäck, G. (2013). The development of pointing perception in infancy: Effects of communicative signals on covert shifts of attention. *Developmental Psychology*, **49** (10), 1898-1908.

Downing, P. E., Dodds, C. M., & Bray, D. (2004). Why does the gaze of others direct visual attention. *Visual Cognition*, **11** (1), 71-79.

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Visual Cognition*, **6** (5), 509-540.

Efran, J.S. (1968). Looking for approval: effect on visual behavior of approbation from persons differing in importance, *Journal of Personality and Social Psychology*, **10**, 21-25.

Efran , J. S., Broughton , A. (1966). Effect of expectancies for social approval on visual behavior. *Journal of Personality and Social Psychology*, **4**, 103-107.

Ellyson, S.L., Davidio, J.F., & Fehr, B.J. (1981). Visual behavior and dominance in women and men. In C. Mayo and N.M. Henley (Eds), *Gender and Nonverbal Behavior*, New York: Springer-Verlag.

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioural Reviews*, **24**, 581-604.

Exline, R. V., Ellyson, S. L., & Long, B. (1975). Visual behavior as an aspect of power role relationships. In P. Pliner, L. Krames, & T. Alloway (Eds.), *Nonverbal communication of aggression* (pp. 53–759). New York: Plenum Press.

Exline, R.V., & Winters, L.C. (1966). Affective relations and mutual glances in dyads. In S.S. Tomkins and C.Izard (Eds), *Affect, Cognition and Personality*, London: Tavistock.

Foulsham, T., Cheng, J.T., Tracy, J.L., Henrich, J. & Kingstone, A. (2010). Gaze allocation in a dynamic situation: Effects of social status and speaking. *Cognition*, **117**, 319-331.

Fox, J., & Weisberg, S. (2011). *An {R} Companion to Applied Regression, Second Edition.* Thousand Oaks, CA: Sage.

Freeth, M., Foulsham, T., & Kingstone, A. (2013). What affects social attention? Social presence, eye contact and autistic traits. *PLoS One*, **8** (1), e53286.

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin and Review*, **5**, 490-495.

Friesen, C. K. Ristic, J., & Kingstone, A. (2004). Attentional effects of counterpredictive gaze and arrow cues. *Journal of Experimental Psychology: Human Perception and Performance*, **30** (2), 319-329.

Gallup, A. C., Chong, A., & Couzin, I. D. (2012). The directional flow of visual information transfer between pedestrians. *Biology Letters*, **8** (4), 520-522.

Gallup, A. C., Hale, J. J., Sumpter, D. J. T., Garnier, S., Kacelnik, A., Krebs, J. R., & Couzin, I. D. (2012). Visual attention and the acquisition of information in human crowds. *Psychological and Cognitive Sciences*, **109** (19), 7245-7250.

Gibson, B. S., & Kingstone, A. (2006). Visual attention and the semantics of space: Beyond central and peripheral cues. *Psychological Science*, ***17***, 622–627.

Gullberg, M. (2002). Eye Movements and Gestures in Human Interaction. In Hyona, J., Radach, R., & Deubel, H. (Eds.): *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research* (pp. 685-703). Amsterdam, The Netherlands: Elsevier.

Hanna, J. E. & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, **57**, 596-615.

Hill, J. L., Patel, S., Gu, X., Seyedali, N. S., Bachevalier, J., & Sereno, A. B. (2010). Social orienting: Reflexive versus voluntary control. *Vision Research*, **50**, 2080-2092.

Hood, B.M., Willen, J.D., Driver, J. (1998). Adult's Eyes Trigger Shifts of Visual Attention in Human Infants. *Psychological Science*. **9**(2), 131-134.

Hothorn, T., Bretz, F., & Westfall, P., (2008). Simultaneous Inference in General Parametric Models, *Biometrical Journal*, **50** (3), 346--363.

Itier, R.J., Villate, C., & Ryan, J.D. (2007). Eyes always attract attention but gaze orienting is task-dependent: Evidence from eye movement monitoring. *Neuropsychologia*, **45**, 1019-1028.

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., Koch, I. (2010). Control and interference in task switching – A review, *Psychological Bulletin*, **136** (5), 849-874

Kingstone, A., Friesen, C. K., & Gazzaniga, M. S. (2000). Reflexive joint attention depends on lateralized cortical connections. *Psychological Science*, **11** (2), 159-166.

Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive Ethology: A new approach for studying human cognition. *British Journal of Psychology*, **99**, 317–340.

Kleck, R.E., & Nuessle, W. (1968). Congruence between the Indicative and Communicative Functions of Eye-Contact in Interpersonal Relations. *British Journal of Social and Clinical Psychology*, **7**, 241-246.

Knoeferle, P., & Kreysa, H. (2012). Can speaker gaze modulate syntactic structuring and thematic role assignment during spoken sentence comprehension? *Frontiers in Psychology*, **3**, 538.

Koch. I., Gade, M., Schuch, S., & Philipp, A. M. (2010). The role of inhibition in task switching: A review, *Psychonomic Bulletin & Review*, **17** (1), 1-14.

Kuhn, G., Benson, V. (2007). The influence of eye-gaze and arrow pointing distractor cues on voluntary eye movements. *Perception and Psychophysics*, **69** (6), 966-971.

Kuhn, G., Kingstone, A. (2009). Look away! Eyes and arrows engage oculomotor responses automatically. *Attention, Perception and Psychophysics*, **71** (2), 314-327.

LaFrance, M., & Mayo, C. (1976). Racial difference in gaze behavior during conversations: two systematic observational studies, *Journal of Personality and Social Psychology*, **33**, 547-552.

Lai, H., Li, F., & Wechsler. (2007). Robust face recognition strategies using feed-forward architectures and parts. *Analysis and modeling of faces and gestures*. Springer: Berlin.

Laidlaw, K. E. W., Foulsham, T., Kuhn, G., & Kingstone, A. (2011). Potential social interactions are important to social attention. *Proceedings of the National Academy of Sciences*, **108**, 5548-5553.

Laidlaw, K. E. W., Risko, E. F., & Kingstone, A. (2012). A new look at social attention: Orienting to the eyes is not (entirely) under volitional *control*. *Journal of Experimental Psychology: Human Perception & Performance*, **38** (5), 1132-1143.

Langdon, R., & Smith, P. (2005) Spatial cueing by social versus nonsocial directional signals, *Visual Cognition*, **12** (8), 1497-1527.

Langton, S. R. H., & Bruce, V. (1999) Reflexive visual orienting in response to the social attention of others, *Visual Cognition*, **6** (5), 541-567.

Levy, J. Foulsham, T., Kingstone, A. (2013). Monsters are people too, *Biology Letters*, **9**, 1-4.

Liuzza, M.T., Cazzato, V., Vecchione, M., Crostella, F., Caprara, G. V., & Aglioti, S.M. (2011). Follow my eyes: The gaze of politicians reflexively captures the gaze of ingroup voters. *PloS ONE,* **6** (9), e25117.

Lorincz, E. N.. Baker, C, I., Perrett, D. I. (1999). Visual cues for attention following in rhesus monkeys. *Current Psychology of Cognition*, **18**, 973-1003.

Macdonald, R. G., & Tatler, B. W. (2013). Do as eye say: Gaze-cueing and language in a real-world social interaction. *Journal of Vision,* 13 (**4**):6, 1-12.

Mehrabian, A. (1968). Relationship of attitude to seated posture, orientation and distance, *Journal of Personality and Social Psychology*, **10** (1), 26-30.

Meyer, A. S., & Bock, K. (1999). Representations and processes in the production of pronouns: Some perspectives from Dutch. *Journal of Memory and Language*, **41** (2), 281-301.

Milgram, S., Bickman, L., & Berkowitz, L. (1969) Note on the drawing power of crowds of different size. *Journal of Personality and Social Psychology*, **13**, 79–82.

Modigliani, A. (1971). Embarrassment, facework and eye contact: Testing a theory of embarrassment, *Journal of Personality and Social Psychology*, **17**(1), 15-24.

Nagy, A. L., Sanchez, R. R., & Hughes, T. C. (1990). Visual search for color differences with foveal and peripheral vision, *Journal of the Optical Society of America,* A**7**, 1995-2001

Nappa, R., & Arnold, J. E. (2014). The road to understanding is paved with the speaker's intentions: Cues to the speaker's attention and intentions affect pronoun comprehension, *Cognitive Psychology*, **70**, 58-81.

Nappa, R., Wessel, A., McEldoon, K. L., Gleitman, L. R., & Trueswell, J. C. (2009). Use of speaker's gaze and syntax in verb learning. *Language Learning and Development*, **5**, 203-234.

Piaget, J. (1972). *The Principles of Genetic Epistemology*. London: Routledge Kegan & Paul.

Pickering, M.J., & Garrod, S. (2004). The interactive-alignment model: Developments and refinements. *Behavioral and Brain Sciences*, **27**, 212-225.

Posner, M.I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, **32**, 3-25.

Quadflieg, S., Mason, M. F., & Macrae. (2004). The owl and the pussycat: Gaze cues and visuospatial orienting, *Psychonomic Bulletic & Review*, **11** (5), 826-831.

Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination: Common ground and the coupling of eye movements during dialogue. *Psychological Science*, **18** (5), 407-413.

Ricciardelli, P., Carcagno, S., Vallar, G., & Bricolo, E. (2013). Is gaze following purely reflexive or goal-directed instead? Revisiting the automaticity of orienting attention by gaze cues. *Experimental Brain Research,* **224** (1), 93-106.

Ricciardelli, P., Bricolo, E., Aglioti, S. M., & Chelazzi, L. (2002). My eyes want to look where your eyes are looking: Exploring the tendency to imitate another individual's gaze. *Neuroreport*, **13** (17), 2259-2264.

Ristic, K., Friesen, C. K., & Kingstone, A. (2002). Are eyes special? It depends on how you look at it. *Psychonomic Bulletin and Review*, **9** (3), 507-513.

Ristic, K., & Kingstone, A. (2005). Taking control of reflexive social attention *Cognition*, **94**, B55-B65.

Scaife, M., & Bruner, J. (1975). The capacity for joint visual attention in the infant. *Nature*, **253**, 265–266.

Schmitt, B. M., Meyer, A. S., & Levelt, W. J. M. (1999). Lexical access in the production of pronouns. *Cognition*, **69** (3), 313-335.

Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology*, **18**, 668-671.

Smilek, D., Birmingham, E., Cameron, D., Bischof, W., & Kingstone, A. (2006). Cognitive Ethology and exploring attention in real-world scenes. *Brain Research*, **1080** (1), 101-119.

Staudte, M., & Crocker, M. W. (2011). Investigating joint attention mechanisms through spoken human-robot interaction, *Cognition*, **120**, 268-291.

Staudte, M., Crocker, M. W., Heloir, A., & Kipp, M. (2014). The influence of speaker gaze on listener comprehension: Contrasting visual versus intentional accounts", *Cognition*, **133** (1), 317-328.

Tipples, J. (2002). Eye gaze is not unique: automatic orienting in response to uninformative arrows. *Psychonomic Bulletin & Review* **9**, 314-318.

Tomasello, M., Hare, B., Lehmann, H. & Call, J. (2007). Reliance on head versus eyes in the gaze following of great apes and human infants: The cooperative eye hypothesis. *Journal of Human Evolution*, **52***,* 314-320.

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, **7** (14):4, 1-17.

Tomkins, S.S., & McCarter, R. (1964). What and Where Are the Primary Affects? Some Evidence for a Theory. *Perceptual and Motor Skills*, **18**, 119-158.

Watson, O.M., & Graves, T.D. (1966). Quantitative research in proxemics behavior, *American Anthropologist*, **68** (4), 971—985.

Yarbus, A.L. (1967). *Eye Movements and Vision.* New York: Plenum Press.

Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PLoS ONE*, **8** (11), e79659.

Zwickel, J., & Võ, M. L.-H. (2010). How the Presence of Humans Biases Gaze Movements. *Psychonomic Bulletin & Review*, **17** (2), 257-262.

Appendix 1: Simplifying LMMs

Following the guidance of Barr, Levy, Scheepers and Tily (2013), the linear mixed effects models in this thesis use the maximal (most complex) random effects structure. These random slope and intercept models consider the different effects that the fixed factors have on different subjects. However, these maximal models can often fail to converge if there are not enough observations in the data and this was often the case in the present thesis. When maximal models failed to converge, simpler models were attempted. Models were simplified, when necessary, by first removing correlations between random-slopes and intercepts. After this, the slope of any interaction from the random factor was removed. If the model still failed to converge, the intercept was removed, followed by a fixed factor slope. The most complex model that successfully converged was always used.

This appendix provides a comprehensive outline of how models were simplified throughout this thesis for models with one or two fixed factors and one or two random factors.

Simplifying models with one fixed and one random factor

These models are the easiest to simplify, as they contain only one fixed and random factor, meaning that there is no need for a predetermined hierarchy of factors. Below are all possible model structures of measure *x* for fixed factor *a* and random factor *Subject (S)* in order of desirability. If the first structure did not convert, the next would be attempted until a model that did not fail to converge was found. The “1” in a random effects structure represents the intercept and the use of "||" represents the absence of correlation parameters between random intercepts and slopes being used in the model.

1) $x \sim a + \ (1 + a|S)$ 2) $x \sim a + \ (1 + a||S)$

3) $x \sim a + \ (0 + a|S)$ 4) $x \sim a + \ (1|S)$

Simplifying models with one fixed and two random factors

In all models with two random factors the *Item* factor was the first to have its structure simplified. Below are all possible model structures of measure *x* for fixed factor *a* and random factors *Subject (S)* and *Item (I)* in order of desirability. If the first structure did not convert, the next would be attempted until a model that did not fail to converge was found.

1) $x \sim a + (1 + a|S) + (1 + a|I)$ 2) $x \sim a + (1 + a|S) + (1 + a||I)$

3) $x \sim a + (1 + a|S) + (0 + a|I)$ 4) $x \sim a + (1 + a|S) + (1|I)$

5) $x \sim a + (1 + a||S) + (1 + a|I)$ 6) $x \sim a + (1 + a||S) + (1 + a||I)$

7) $x \sim a + (1 + a||S) + (0 + a|I)$ 8) $x \sim a + (1 + a||S) + (1|I)$

9) $x \sim a + (0 + a|S) + (1 + a|I)$ 10) $x \sim a + (0 + a|S) + (1 + a||I)$

11) $x \sim a + (0 + a|S) + (0 + a|I)$ 12) $x \sim a + (0 + a|S) + (1|I)$

13) $x \sim a + (1|S) + (1 + a|I)$

14) $x \sim a + (1|S) + (1 + a||I)$

15) $x \sim a + (1|S) + (0 + a|I)$

16) $x \sim a + (1|S) + (1|I)$

Simplifying models with two fixed factor and one random factor

In all models with two fixed factors one of the fixed-factor slopes (*b*) was removed first. Below are all possible model structures of measure *x* for fixed factors *a* and *b* and random factor *Subject (S)*. If the first structure did not converge, the next would be attempted until a model that did not fail to converge was found.

1) $x \sim a*b + (1+a*b|S)$

2) $x \sim a*b + (1+a*b||S)$

3) $x \sim a*b + (1+a+b||S)$

4) $x \sim a*b + (0+a+b||S)$

5) $x \sim a*b + (1+a||S)$

6) $x \sim a*b + (0+a|S)$

7) $x \sim a*b + (1+b||S)$

8) $x \sim a*b + (0+b|S)$

9) $x \sim a*b + (1|S)$

Simplifying models with two fixed factor and two random factors

In these types of model there are a number of different ways to simplify the random effects structure. In this thesis, *Item* was simplified first and in both random effects structures, one of the fixed-factor slopes (*b*) was consistently removed first. Below are all possible model structures of measure *x* for fixed factors *a* and *b* and random factors *Subject (S)* and *Item (I)*. If the first structure did not converge, the next would be attempted until a model that did not fail to converge was found.

1) x ~ a*b + (1+a*b|S) + (1+a*b|I)

2) x ~ a*b + (1+a*b|S) + (1+a*b||I)

3) x ~ a*b + (1+a*b|S) + (1+a+b|I)

4) x ~ a*b + (1+a*b|S) + (0+a+b||I)

5) x ~ a*b + (1+a*b|S) + (1+a||I)

6) x ~ a*b + (1+a*b|S) + (0+a|I)

7) x ~ a*b + (1+a*b|S) + (1+b||I)

8) x ~ a*b + (1+a*b|S) + (0+b|I)

9) x ~ a*b + (1+a*b|S) + (1|I)

10) x ~ a*b + (1+a*b||S) + (1+a*b|I)

11) x ~ a*b + (1+a*b||S) + (1+a*b||I)

12) x ~ a*b + (1+a*b||S) + (1+a+b||I)

13) x ~ a*b + (1+a*b||S) + (0+a+b||I)

14) x ~ a*b + (1+a*b||S) + (1+a||I)

15) x ~ a*b + (1+a*b||S) + (0+a|I)

16) x ~ a*b + (1+a*b||S) + (1+b||I)

17) x ~ a*b + (1+a*b||S) + (0+b|I)

18) x ~ a*b + (1+a*b||S) + (1|I)

19) x ~ a*b + (1+a+b||S) + (1+a*b|I)

20) x ~ a*b + (1+a+b||S) + (1+a*b||I)

21) x ~ a*b + (1+a+b||S) + (1+a+b||I)

22) x ~ a*b + (1+a+b||S) + (0+a+b||I)

23) x ~ a*b + (1+a+b||S) + (1+a||I)

24) x ~ a*b + (1+a+b||S) + (0+a|I)

25) x ~ a*b + (1+a+b||S) + (1+b||I)

26) x ~ a*b + (1+a+b||S) + (0+b|I)

27) x ~ a*b + (1+a+b||S) + (1|I)

28) x ~ a*b + (0+a+b||S) + (1+a*b|I)

29) x ~ a*b + (0+a+b||S) + (1+a*b||I)

30) x ~ a*b + (0+a+b||S) + (1+a+b||I)

31) x ~ a*b + (0+a+b||S) + (0+a+b||I)

32) x ~ a*b + (0+a+b||S) + (1+a||I)

33) x ~ a*b + (0+a+b||S) + (0+a|I)

34) x ~ a*b + (0+a+b||S) + (1+b||I)

35) x ~ a*b + (0+a+b||S) + (0+b|I)

36) x ~ a*b + (0+a+b||S) + (1|I)

37) x ~ a*b + (1+a||S) + (1+a*b|I)

38) x ~ a*b + (1+a||S) + (1+a*b||I)

39) x ~ a*b + (1+a||S) + (1+a+b|I)

40) x ~ a*b + (1+a||S) + (0+a+b||I)

41) x ~ a*b + (1+a||S) + (1+a||I)

42) x ~ a*b + (1+a||S) + (0+a|I)

43) x ~ a*b + (1+a||S) + (1+b||I)

44) x ~ a*b + (1+a||S) + (0+b|I)

45) x ~ a*b + (1+a||S) + (1|I)

46) x ~ a*b + (0+a|S) + (1+a*b|I)

47) x ~ a*b + (0+a|S) + (1+a*b||I)

48) x ~ a*b + (0+a|S) + (1+a+b||I)

49) x ~ a*b + (0+a|S) + (0+a+b||I)

50) x ~ a*b + (0+a|S) + (1+a||I)

51) x ~ a*b + (0+a|S) + (0+a|I)

52) x ~ a*b + (0+a|S) + (1+b||I)

53) x ~ a*b + (0+a|S) + (0+b|I)

54) x ~ a*b + (0+a|S) + (1|I)

55) x ~ a*b + (1+b||S) + (1+a*b|I)

56) x ~ a*b + (1+b||S) + (1+a*b||I)

57) x ~ a*b + (1+b||S) + (1+a+b||I)

58) x ~ a*b + (1+b||S) + (0+a+b||I)

59) x ~ a*b + (1+b||S) + (1+a||I)

60) x ~ a*b + (1+b||S) + (0+a|I)

61) x ~ a*b + (1+b||S) + (1+b||I)

62) x ~ a*b + (1+b||S) + (0+b|I)

63) x ~ a*b + (1+b||S) + (1|I)

64) x ~ a*b + (0+b|S) + (1+a*b|I)

65) x ~ a*b + (0+b|S) + (1+a*b||I)

66) x ~ a*b + (0+b|S) + (1+a+b||I)

67) x ~ a*b + (0+b|S) + (0+a+b||I)

68) x ~ a*b + (0+b|S) + (1+a||I)

69) x ~ a*b + (0+b|S) + (0+a|I)

70) x ~ a*b + (0+b|S) + (1+b||I)

71) x ~ a*b + (0+b|S) + (0+b|I)

72) x ~ a*b + (0+b|S) + (1|I)

73) $x \sim a*b + (1|S) + (1+a*b|I)$

74) $x \sim a*b + (1|S) + (1+a*b||I)$

75) $x \sim a*b + (1|S) + (1+a+b|\mathrm{I})$

76) $x \sim a*b + (1|S) + (0+a+b||I)$

77) $x \sim a*b + (1|S) + (1+a||I)$

78) $x \sim a*b + (1|S) + (0+a|I)$

79) $x \sim a*b + (1|S) + (1+b||I)$

80) $x \sim a*b + (1|S) + (0+b|I)$

81) $x \sim a*b + (1|S) + (1|I)$

Appendix 2: R-script for Linear Mixed Effects Models used in thesis

Chapter 2

2.1)

#sum coding

data_shared$roles.roles <- ifelse(data_shared$roles == "roles", 1, 0)

meanroles <- mean(data_shared$roles.roles)

data_shared$roles.sum <- data_shared$roles.roles-meanroles

#model

model <- glmer(initiator3 ~ roles.sum + (roles.sum||video), family = binomial, data = data_shared)

2.2)

#sum coding

data_shared3$roles.roles <- ifelse(data_shared3$roles == "roles", 1, 0)

meanroles <- mean(data_shared3$roles.roles)

data_shared3$roles.sum <- data_shared3$roles.roles-meanroles

data_shared3$initiatorA <- ifelse(data_shared3$initiator == "a", 1, 0)

meaninitiator <- mean(data_shared3$initiatorA)

data_shared3$initiator.sum <- data_shared3$initiatorA-meaninitiator

#model

```
model <- lmer(gap.log ~ roles.sum*initiator.sum + (roles.sum*initiator.sum||video), data = data_shared3)
```

2.3)

#sum coding

```
data_mutual2$roles.roles <- ifelse(data_mutual2$roles == "roles", 1, 0)

meanroles <- mean(data_mutual2$roles.roles)

data_mutual2$roles.sum <- data_mutual2$roles.roles-meanroles


data_mutual2$person.person <- ifelse(data_mutual2$person == "a", 1, 0)

meanperson <- mean(data_mutual2$person.person)

data_mutual2$person.sum <- data_mutual2$person.person-meanperson
```

#model

```
model <- lmer(time.log ~ roles.sum*person.sum + (roles.sum*person.sum||video), data = data_mutual2)
```

2.4)

#sum coding

```
data_instruction$roles.roles <- ifelse(data_instruction$roles == "roles", 1, 0)

meanroles <- mean(data_instruction$roles.roles)

data_instruction$roles.sum <- data_instruction$roles.roles-meanroles
```

#model

```
model <- glmer(look ~ roles.sum + (roles.sum|video), family = binomial, data= data_instruction)
```

2.5)

#sum coding

data.instructions.look$roles.roles <- ifelse(data.instructions.look$roles == "roles", 1, 0)

meanroles <- mean(data.instructions.look$roles.roles)

data.instructions.look$roles.sum <- data.instructions.look$roles.roles-meanroles

model <- glmer(who ~ roles.sum + (roles.sum|video), family = binomial, data= data.instructions.look)

2.6)

#sum coding

data_reference_object5$roles.roles <- ifelse(data_reference_object5$roles == "roles", 1, ifelse(data_reference_object5$roles == "no roles", 0, NA))

meanroles <- mean(data_reference_object5$roles.roles, na.rm = TRUE)

data_reference_object5$roles.sum <- data_reference_object5$roles.roles-meanroles

data_reference_object5$person.person <- ifelse(data_reference_object5$person == "a", 1, 0)

meanperson <- mean(data_reference_object5$person.person, na.rm = TRUE)

data_reference_object5$person.sum <- data_reference_object5$person.person-meanperson

 # model

model <- lmer(look.minus.start.log ~ roles.sum*person.sum + (roles.sum*person.sum||video), data = data_reference_object5)

2.7)

#Sum coding

meannoun <- mean(data_reference3$proper.noun, na.rm = TRUE)

data_reference3$noun.sum <- data_reference3$proper.noun-meannoun

```
data_reference3$roles.roles <- ifelse(data_reference3$roles == "roles", 1,
ifelse(data_reference3$roles == "no roles", 0, NA))

meanroles <- mean(data_reference3$roles.roles, na.rm = TRUE)

data_reference3$roles.sum <- data_reference3$roles.roles-meanroles
```

#model

```
model <- lmer(alook.minus.start.log ~ roles.sum*noun.sum +
    (roles.sum*noun.sum||video), data = data_reference3)
```

2.8)

#sum coding

```
meannoun <- mean(data_reference3$proper.noun, na.rm = TRUE)

data_reference3$noun.sum <- data_reference3$proper.noun-meannoun


data_reference3$roles.roles <- ifelse(data_reference3$roles == "roles", 1,
ifelse(data_reference3$roles == "no roles", 0, NA))

meanroles <- mean(data_reference3$roles.roles, na.rm = TRUE)

data_reference3$roles.sum <- data_reference3$roles.roles-meanroles
```

#model

```
model <- lmer(blook.minus.start.log ~ roles.sum*noun.sum +
    (roles.sum*noun.sum||video), data = data_reference3)
```

2.9)

#Sum coding

```
meanspatial <- mean(data_reference3$spatial, na.rm = TRUE)

data_reference3$spatial.sum <- data_reference3$spatial-meanspatial


data_reference3$roles.roles <- ifelse(data_reference3$roles == "roles", 1,
ifelse(data_reference3$roles == "no roles", 0, NA))

meanroles <- mean(data_reference3$roles.roles, na.rm = TRUE)

data_reference3$roles.sum <- data_reference3$roles.roles-meanroles
```

#model

```
model <- lmer(alook.minus.start.log ~ roles.sum*spatial.sum +
        (roles.sum*spatial.sum||video), data = data_reference3)
```

2.10)

#sum coding

```
meanspatial <- mean(data_reference3$spatial, na.rm = TRUE)

data_reference3$spatial.sum <- data_reference3$spatial-meanspatial


data_reference3$roles.roles <- ifelse(data_reference3$roles == "roles", 1,
ifelse(data_reference3$roles == "no roles", 0, NA))

meanroles <- mean(data_reference3$roles.roles, na.rm = TRUE)

data_reference3$roles.sum <- data_reference3$roles.roles-meanroles
```

#model

```
model <- lmer(blook.minus.start.log ~ roles.sum*spatial.sum +
        (roles.sum*spatial.sum||video), data = data_reference3)
```

2.11)

#sum coding

```
meannoun <- mean(data_reference_last$proper.noun, na.rm = TRUE)

data_reference_last$noun.sum <- data_reference_last$proper.noun-meannoun


data_reference_last$roles.roles <- ifelse(data_reference_last$roles == "roles", 1,
ifelse(data_reference_last$roles == "no roles", 0, NA))

meanroles <- mean(data_reference_last$roles.roles, na.rm = TRUE)

data_reference_last$roles.sum <- data_reference_last$roles.roles-meanroles
```

#model

```
model <- glmer(spatial ~ noun.sum + (noun.sum||video), family = binomial, data =
data_reference_last)
```

CHAPTER 3

3.1)

#model

```
model.output3 = glmer(start.with.look.to.me. ~ gaze.condition +  (1|ParticipantF),
family="binomial", data = dataready)
```

3.2)

#sum coding

```
dataready3$AQgroupEFF <- ifelse(dataready3$groupalt== "1", 0,
ifelse(dataready3$groupalt== "2", 1, NA))

mean.group <- mean(dataready3$AQgroupEFF, na.rm = T)

dataready3$AQ.sum <- ifelse(dataready3$groupalt== "1", 0-mean.group,
ifelse(dataready3$groupalt== "2", 1-mean.group, NA))
```

#model

```
model.output = glmer(start.with.look.to.me. ~ AQ.sum +  (1|ParticipantF), family =
binomial, data = dataready3)
```

3.3)

#model

```
model.output3 = glmer(look.at.point.of.gaze ~ gaze.condition +  (1|ParticipantF)+
(1|object), family="binomial", data = dataready)
```

3.4, 3.5, 3.6)

#sum coding

```
dataready3$AQgroupEFF <- ifelse(dataready3$groupalt== "1", 0,
ifelse(dataready3$groupalt== "2", 1, NA))

mean.group <- mean(dataready3$AQgroupEFF, na.rm = T)

dataready3$AQ.sum <- ifelse(dataready3$groupalt== "1", 0-mean.group,
ifelse(dataready3$groupalt== "2", 1-mean.group, NA))
```

#model

```
model.output = glmer(look.at.point.of.gaze ~ AQ.sum +  (1|ParticipantF) +
(AQ.sum|object), family = binomial, data = dataready3)

summary(model.output)
```

3.7)

#sum coding

```
dataready3 <- dataready3[!(dataready3$gaze.condition == "AG"),]

dataready3$gaze.conditionEFF <- ifelse(dataready3$gaze.condition == "NG", 0,
ifelse(dataready3$gaze.condition == "G", 1, NA))

mean.gaze.GNG <- mean(dataready3$gaze.conditionEFF, na.rm = T)

dataready3$gaze.condition.sum <- ifelse(dataready3$gaze.condition == "NG", 0-
mean.gaze.GNG, ifelse(dataready3$gaze.condition == "G", 1-mean.gaze.GNG, NA))
```

dataready3$featural.or.spatialEFF <- ifelse(dataready3$featural.or.spatial== "featural", 0, ifelse(dataready3$featural.or.spatial== "spatial", 1, NA))

mean.lang <- mean(dataready3$featural.or.spatialEFF, na.rm = T)

dataready3$featural.or.spatial.sum <- ifelse(dataready3$featural.or.spatial== "featural", 0-mean.lang, ifelse(dataready3$featural.or.spatial== "spatial", 1-mean.lang, NA))

#Model

model.output = lmer(prop.time.looking ~ gaze.condition.sum*featural.or.spatial.sum+ (featural.or.spatial.sum|ParticipantF) + (gaze.condition.sum*featural.or.spatial.sum|object), data = dataready3)

3.8)

#sum coding

dataready3 <- dataready3[!(dataready3$gaze.condition == "G"),]

dataready3$gaze.conditionEFF <- ifelse(dataready3$gaze.condition == "NG", 0, ifelse(dataready3$gaze.condition == "AG", 1, NA))

mean.gaze.GNG <- mean(dataready3$gaze.conditionEFF, na.rm = T)

dataready3$gaze.condition.sum <- ifelse(dataready3$gaze.condition == "NG", 0-mean.gaze.GNG, ifelse(dataready3$gaze.condition == "AG", 1-mean.gaze.GNG, NA))

dataready3$featural.or.spatialEFF <- ifelse(dataready3$featural.or.spatial== "featural", 0, ifelse(dataready3$featural.or.spatial== "spatial", 1, NA))

mean.lang <- mean(dataready3$featural.or.spatialEFF, na.rm = T)

dataready3$featural.or.spatial.sum <- ifelse(dataready3$featural.or.spatial== "featural", 0-mean.lang, ifelse(dataready3$featural.or.spatial== "spatial", 1-mean.lang, NA))

#Model

model.output = lmer(prop.time.looking ~ gaze.condition.sum*featural.or.spatial.sum+ (featural.or.spatial.sum|ParticipantF) + (gaze.condition.sum*featural.or.spatial.sum|object), data = dataready3)

3.9)

#sum coding

dataready3 <- dataready3[dataready3$gaze.condition == "G",]

dataready3$featural.or.spatialEFF <- ifelse(dataready3$featural.or.spatial== "featural", 0, ifelse(dataready3$featural.or.spatial== "spatial", 1, NA))

mean.lang <- mean(dataready3$featural.or.spatialEFF, na.rm = T)

dataready3$featural.or.spatial.sum <- ifelse(dataready3$featural.or.spatial== "featural", 0-mean.lang, ifelse(dataready3$featural.or.spatial== "spatial", 1-mean.lang, NA))

dataready3$AQgroupEFF <- ifelse(dataready3$groupalt== "1", 0, ifelse(dataready3$groupalt== "2", 1, NA))

mean.group <- mean(dataready3$AQgroupEFF, na.rm = T)

dataready3$AQ.sum <- ifelse(dataready3$groupalt== "1", 0-mean.group, ifelse(dataready3$groupalt== "2", 1-mean.group, NA))

#model

model.output = lmer(prop.time.looking ~ featural.or.spatial.sum*AQ.sum + (featural.or.spatial.sum|ParticipantF) + (featural.or.spatial.sum*AQ.sum|object), data = dataready3)

3.10)

#sum coding

```
dataready3 <- dataready3[dataready3$gaze.condition != "G",]

dataready3$AQgroupEFF <- ifelse(dataready3$groupalt== "1", 0,
ifelse(dataready3$groupalt== "2", 1, NA))

mean.group <- mean(dataready3$AQgroupEFF, na.rm = T)

dataready3$AQ.sum <- ifelse(dataready3$groupalt== "1", 0-mean.group,
ifelse(dataready3$groupalt== "2", 1-mean.group, NA))
```

#model

```
model.output = lmer(prop.time.looking ~ AQ.sum +  (1|ParticipantF) +
(AQ.sum|object), data = dataready3)
```

3.11)

#sum coding

```
dataready3 <- dataready[!(dataready$gaze.condition == "AG"),]

dataready3$gaze.conditionEFF <- ifelse(dataready3$gaze.condition == "NG", 0,
ifelse(dataready3$gaze.condition == "G", 1, NA))

mean.gaze.GNG <- mean(dataready3$gaze.conditionEFF, na.rm = T)

dataready3$gaze.condition.sum <- ifelse(dataready3$gaze.condition == "NG", 0-
mean.gaze.GNG, ifelse(dataready3$gaze.condition == "G", 1-mean.gaze.GNG, NA))

dataready3$featural.or.spatialEFF <- ifelse(dataready3$featural.or.spatial==
"featural", 0, ifelse(dataready3$featural.or.spatial== "spatial", 1, NA))

mean.lang <- mean(dataready3$featural.or.spatialEFF, na.rm = T)

dataready3$featural.or.spatial.sum <- ifelse(dataready3$featural.or.spatial==
"featural", 0-mean.lang, ifelse(dataready3$featural.or.spatial== "spatial", 1-
mean.lang, NA))
```

#Model

```
model.output3 = glmer(correct.direction ~ gaze.condition.sum*featural.or.spatial.sum
    + (featural.or.spatial.sum|ParticipantF) + (0+featural.or.spatial.sum|object),
    family="binomial", data = dataready3)
```

3.12)

#sum coding

```
dataready3 <- dataready[!(dataready$gaze.condition == "G"),]

dataready3$gaze.conditionEFF <- ifelse(dataready3$gaze.condition == "NG", 0,
ifelse(dataready3$gaze.condition == "AG", 1, NA))

mean.gaze.GNG <- mean(dataready3$gaze.conditionEFF, na.rm = T)

dataready3$gaze.condition.sum <- ifelse(dataready3$gaze.condition == "NG", 0-
mean.gaze.GNG, ifelse(dataready3$gaze.condition == "AG", 1-mean.gaze.GNG,
NA))

dataready3$featural.or.spatialEFF <- ifelse(dataready3$featural.or.spatial==
"featural", 0, ifelse(dataready3$featural.or.spatial== "spatial", 1, NA))

mean.lang <- mean(dataready3$featural.or.spatialEFF, na.rm = T)

dataready3$featural.or.spatial.sum <- ifelse(dataready3$featural.or.spatial==
"featural", 0-mean.lang, ifelse(dataready3$featural.or.spatial== "spatial", 1-
mean.lang, NA))
```

#Model

```
model.output3 = glmer(correct.direction ~ gaze.condition.sum*featural.or.spatial.sum
    + (gaze.condition.sum|ParticipantF) + (0+ gaze.condition.sum
    +featural.or.spatial.sum|object), family="binomial", data = dataready3)
```

3.13)

#sum coding

```
dataready2 <- dataready[dataready$look.at.point.of.gaze == "1", ]
```

```
dataready3 <- dataready2[!(dataready2$gaze.condition == "AG"),]


dataready3$gaze.conditionEFF <- ifelse(dataready3$gaze.condition == "NG", 0,
ifelse(dataready3$gaze.condition == "G", 1, NA))

mean.gaze.GNG <- mean(dataready3$gaze.conditionEFF, na.rm = T)

dataready3$gaze.condition.sum <- ifelse(dataready3$gaze.condition == "NG", 0-
mean.gaze.GNG, ifelse(dataready3$gaze.condition == "G", 1-mean.gaze.GNG, NA))


dataready3$featural.or.spatialEFF <- ifelse(dataready3$featural.or.spatial==
"featural", 0, ifelse(dataready3$featural.or.spatial== "spatial", 1, NA))

mean.lang <- mean(dataready3$featural.or.spatialEFF, na.rm = T)

dataready3$featural.or.spatial.sum <- ifelse(dataready3$featural.or.spatial==
"featural", 0-mean.lang, ifelse(dataready3$featural.or.spatial== "spatial", 1-
mean.lang, NA))
```

#model

```
model.output3 = glmer(correct.direction ~ gaze.condition.sum*featural.or.spatial.sum
+ (featural.or.spatial.sum|ParticipantF) +
(0+featural.or.spatial.sum+gaze.condition.sum||object), family="binomial", data =
dataready3)
```

3.14)

#sum coding

```
dataready2 <- dataready[dataready$look.at.point.of.gaze == "1", ]


dataready3 <- dataready2[!(dataready2$gaze.condition == "G"),]


dataready3$gaze.conditionEFF <- ifelse(dataready3$gaze.condition == "NG", 0,
ifelse(dataready3$gaze.condition == "AG", 1, NA))

mean.gaze.GNG <- mean(dataready3$gaze.conditionEFF, na.rm = T)
```

dataready3$gaze.condition.sum <- ifelse(dataready3$gaze.condition == "NG", 0-mean.gaze.GNG, ifelse(dataready3$gaze.condition == "AG", 1-mean.gaze.GNG, NA))

dataready3$featural.or.spatialEFF <- ifelse(dataready3$featural.or.spatial== "featural", 0, ifelse(dataready3$featural.or.spatial== "spatial", 1, NA))

mean.lang <- mean(dataready3$featural.or.spatialEFF, na.rm = T)

dataready3$featural.or.spatial.sum <- ifelse(dataready3$featural.or.spatial== "featural", 0-mean.lang, ifelse(dataready3$featural.or.spatial== "spatial", 1-mean.lang, NA))

#model

model.output3 = glmer(correct.direction ~ gaze.condition.sum*featural.or.spatial.sum + (featural.or.spatial.sum|ParticipantF) + (featural.or.spatial.sum+gaze.condition.sum||object), family="binomial", data = dataready3)

3.15)

#sum coding

dataready3 <- dataready[dataready$gaze.condition == "AG",]

dataready3$featural.or.spatialEFF <- ifelse(dataready3$featural.or.spatial== "featural", 0, ifelse(dataready3$featural.or.spatial== "spatial", 1, NA))

mean.lang <- mean(dataready3$featural.or.spatialEFF, na.rm = T)

dataready3$featural.or.spatial.sum <- ifelse(dataready3$featural.or.spatial== "featural", 0-mean.lang, ifelse(dataready3$featural.or.spatial== "spatial", 1-mean.lang, NA))

dataready3$look.at.point.of.gazeEFF <- ifelse(dataready3$look.at.point.of.gaze== 0, 0, ifelse(dataready3$look.at.point.of.gaze== 1, 1, NA))

mean.look <- mean(dataready3$look.at.point.of.gaze, na.rm = T)

dataready3$look.at.point.of.gaze.sum <- ifelse(dataready3$look.at.point.of.gaze== 0, 0-mean.look, ifelse(dataready3$look.at.point.of.gaze== 1, 1-mean.look, NA))

#model

```
model.output3 = glmer(correct.direction ~
look.at.point.of.gaze.sum*featural.or.spatial.sum +
(featural.or.spatial.sum|ParticipantF) +
(0+look.at.point.of.gaze.sum+featural.or.spatial.sum||object) , family="binomial",
data = dataready3)
```

3.16)

#sum coding

```
dataready3 <- dataready[dataready$gaze.condition == "G",]

dataready3$AQgroupEFF <- ifelse(dataready3$groupalt== "1", 0,
ifelse(dataready3$groupalt== "2", 1, NA))

mean.group <- mean(dataready3$AQgroupEFF, na.rm = T)

dataready3$AQ.sum <- ifelse(dataready3$groupalt== "1", 0-mean.group,
ifelse(dataready3$groupalt== "2", 1-mean.group, NA))
```

#model

```
model.output = glmer(correct.direction ~ AQ.sum +  (1|ParticipantF) +
(AQ.sum|object), family = binomial, data = dataready3)
```

3.17)

#sum coding

```
dataready3 <- dataready[!dataready$gaze.condition == "G",]

dataready3$AQgroupEFF <- ifelse(dataready3$groupalt== "1", 0,
ifelse(dataready3$groupalt== "2", 1, NA))

mean.group <- mean(dataready3$AQgroupEFF, na.rm = T)

dataready3$AQ.sum <- ifelse(dataready3$groupalt== "1", 0-mean.group,
ifelse(dataready3$groupalt== "2", 1-mean.group, NA))
```

#model

```
model.output = glmer(correct.direction ~ AQ.sum +  (1|ParticipantF) + (AQ.sum|object), family = binomial, data = dataready3)
```

3.18)

#sum coding

```
dataready3 <- dataready3[!(dataready3$gaze.condition == "AG"),]

dataready3$gaze.conditionEFF <- ifelse(dataready3$gaze.condition == "NG", 1, ifelse(dataready3$gaze.condition == "G", 0, NA))

mean.gaze.GNG <- mean(dataready3$gaze.conditionEFF, na.rm = T)

dataready3$gaze.condition.sum <- ifelse(dataready3$gaze.condition == "NG", 1-mean.gaze.GNG, ifelse(dataready3$gaze.condition == "G", 0-mean.gaze.GNG, NA))

dataready3$featural.or.spatialEFF <- ifelse(dataready3$featural.or.spatial== "featural", 0, ifelse(dataready3$featural.or.spatial== "spatial", 1, NA))

mean.lang <- mean(dataready3$featural.or.spatialEFF, na.rm = T)

dataready3$featural.or.spatial.sum <- ifelse(dataready3$featural.or.spatial== "featural", 0-mean.lang, ifelse(dataready3$featural.or.spatial== "spatial", 1-mean.lang, NA))
```

#model

```
model.output = lmer(first.look.sentence ~ gaze.condition.sum*featural.or.spatial.sum +  (featural.or.spatial.sum|ParticipantF) + (gaze.condition.sum*featural.or.spatial.sum|object), data = dataready3)
```

3.19)

#sum coding

```
dataready3 <- dataready3[!(dataready3$gaze.condition == "G"),]
```

dataready3$gaze.conditionEFF <- ifelse(dataready3$gaze.condition == "NG", 1, ifelse(dataready3$gaze.condition == "AG", 0, NA))

mean.gaze.GNG <- mean(dataready3$gaze.conditionEFF, na.rm = T)

dataready3$gaze.condition.sum <- ifelse(dataready3$gaze.condition == "NG", 1-mean.gaze.GNG, ifelse(dataready3$gaze.condition == "AG", 0-mean.gaze.GNG, NA))

dataready3$featural.or.spatialEFF <- ifelse(dataready3$featural.or.spatial== "featural", 0, ifelse(dataready3$featural.or.spatial== "spatial", 1, NA))

mean.lang <- mean(dataready3$featural.or.spatialEFF, na.rm = T)

dataready3$featural.or.spatial.sum <- ifelse(dataready3$featural.or.spatial== "featural", 0-mean.lang, ifelse(dataready3$featural.or.spatial== "spatial", 1-mean.lang, NA))

#model

model.output = lmer(first.look.sentence ~ gaze.condition.sum*featural.or.spatial.sum + (featural.or.spatial.sum|ParticipantF) + (gaze.condition.sum*featural.or.spatial.sum|object), data = dataready3)

3.20)

#sum coding

dataready3 <- dataready3[!(dataready3$gaze.condition == "AG"),]

dataready3$gaze.conditionEFF <- ifelse(dataready3$gaze.condition == "NG", 0, ifelse(dataready3$gaze.condition == "G", 1, NA))

mean.gaze.GNG <- mean(dataready3$gaze.conditionEFF, na.rm = T)

dataready3$gaze.condition.sum <- ifelse(dataready3$gaze.condition == "NG", 0-mean.gaze.GNG, ifelse(dataready3$gaze.condition == "G", 1-mean.gaze.GNG, NA))

dataready3$featural.or.spatialEFF <- ifelse(dataready3$featural.or.spatial== "featural", 0, ifelse(dataready3$featural.or.spatial== "spatial", 1, NA))

mean.lang <- mean(dataready3$featural.or.spatialEFF, na.rm = T)

```
dataready3$featural.or.spatial.sum <- ifelse(dataready3$featural.or.spatial==
"featural", 0-mean.lang, ifelse(dataready3$featural.or.spatial== "spatial", 1-
mean.lang, NA))
```

#model

```
model.output = lmer(first.look.sentence ~ gaze.condition.sum*featural.or.spatial.sum
                                        + (featural.or.spatial.sum|ParticipantF) +
          (gaze.condition.sum*featural.or.spatial.sum|object), data = dataready3)
```

3.21)

#sum coding

```
dataready3 <- dataready3[!(dataready3$gaze.condition == "G"),]

dataready3$gaze.conditionEFF <- ifelse(dataready3$gaze.condition == "NG", 0,
ifelse(dataready3$gaze.condition == "AG", 1, NA))

mean.gaze.GNG <- mean(dataready3$gaze.conditionEFF, na.rm = T)

dataready3$gaze.condition.sum <- ifelse(dataready3$gaze.condition == "NG", 0-
mean.gaze.GNG, ifelse(dataready3$gaze.condition == "AG", 1-mean.gaze.GNG,
NA))

dataready3$featural.or.spatialEFF <- ifelse(dataready3$featural.or.spatial==
"featural", 0, ifelse(dataready3$featural.or.spatial== "spatial", 1, NA))

mean.lang <- mean(dataready3$featural.or.spatialEFF, na.rm = T)

dataready3$featural.or.spatial.sum <- ifelse(dataready3$featural.or.spatial==
"featural", 0-mean.lang, ifelse(dataready3$featural.or.spatial== "spatial", 1-
mean.lang, NA))
```

#model

```
model.output = lmer(first.look.sentence ~ gaze.condition.sum*featural.or.spatial.sum
+ (featural.or.spatial.sum|ParticipantF) +
(gaze.condition.sum*featural.or.spatial.sum|object), data = dataready3)
```

3.22)

#sum coding

```
dataready5 <- dataready3[dataready3$gaze.condition == "AG", ]

dataready5$featural.or.spatialEFF <- ifelse(dataready5$featural.or.spatial==
"featural", 0, ifelse(dataready5$featural.or.spatial== "spatial", 1, NA))

mean.lang <- mean(dataready5$featural.or.spatialEFF, na.rm = T)

dataready5$featural.or.spatial.sum <- ifelse(dataready5$featural.or.spatial==
"featural", 0-mean.lang, ifelse(dataready5$featural.or.spatial== "spatial", 1-
mean.lang, NA))

dataready5$look.at.point.of.gazeEFF <- ifelse(dataready5$look.at.point.of.gaze== 0,
0, ifelse(dataready5$look.at.point.of.gaze== 1, 1, NA))

mean.look <- mean(dataready5$look.at.point.of.gaze, na.rm = T)

dataready5$look.at.point.of.gaze.sum <- ifelse(dataready5$look.at.point.of.gaze== 0,
0-mean.look, ifelse(dataready5$look.at.point.of.gaze== 1, 1-mean.look, NA))
```

#model

```
model.output3 = lmer(first.look.sentence ~
look.at.point.of.gaze.sum*featural.or.spatial.sum +
(look.at.point.of.gaze.sum*featural.or.spatial.sum|ParticipantF) +
(look.at.point.of.gaze.sum*featural.or.spatial.sum|object) ,  data = dataready5)
```

3.23, 3.24, 3.25)

#sum coding

```
dataready3$AQgroupEFF <- ifelse(dataready3$groupalt== "1", 0,
ifelse(dataready3$groupalt== "2", 1, NA))

mean.group <- mean(dataready3$AQgroupEFF, na.rm = T)

dataready3$AQ.sum <- ifelse(dataready3$groupalt== "1", 0-mean.group,
ifelse(dataready3$groupalt== "2", 1-mean.group, NA))
```

#model

```
model.output = lmer(first.look.sentence ~ AQ.sum +  (1|ParticipantF) + (
AQ.sum|object), data = dataready3)
```

CHAPTER 4

4.1)

#Sum coding

```
data9_incon$referenceN <- ifelse(data9_incon$reference == 1, 0,
ifelse(data9_incon$reference == 2, 1, 0))

reference_mean <- mean(data9_incon$referenceN)

data9_incon$reference.sum <- ifelse(data9_incon$referenceN == "1", 1-
reference_mean, ifelse(data9_incon$referenceN == "0", 0-reference_mean, NA))
```

#model

```
model.output1 = glmer(followN ~ reference.sum + (reference.sum|subjF) +
(reference.sum|objectF), family = "binomial", data = data9_incon)
```

4.2)

#Sum coding

```
data5_incon$referenceN <- ifelse(data5_incon$reference == 1, 0,
ifelse(data5_incon$reference == 2, 1, 0))

reference_mean <- mean(data5_incon$referenceN)

data5_incon$reference.sum <- ifelse(data5_incon$referenceN == "1", 1-
reference_mean, ifelse(data5_incon$referenceN == "0", 0-reference_mean, NA))
```

#model

```
model.output1 = glmer(followN ~ reference.sum + (reference.sum|subjF) +
(reference.sum|objectF), family = "binomial", data = data5_incon)
```

4.3)

#Sum coding

```
gaze_likelihoodEFF_mean <- mean(data_2_3$gaze_likelihoodEFF)

data_2_3$gaze_likelihood.sum <- ifelse(data_2_3$gaze_likelihoodEFF == "1", 1-
gaze_likelihoodEFF_mean, ifelse(data_2_3$gaze_likelihoodEFF == "0", 0-
gaze_likelihoodEFF_mean, NA))

congruentEEF_mean <- mean(data_2_3$congruentEEF)

data_2_3$congruentEEF.sum <- ifelse(data_2_3$congruentEEF == "1", 1-
congruentEEF_mean, ifelse(data_2_3$congruentEEF == "0", 0-congruentEEF_mean,
NA))
```

#model

```
model.output1 = glmer(correct ~ gaze_likelihood.sum*congruentEEF.sum +
(gaze_likelihood.sum*congruentEEF.sum||subjF) + (0+gaze_likelihood.sum|objectF),
family = "binomial", data = data_2_3)

summary(model.output1)
```

4.4)

#sum coding

language_likelihoodEFF_mean <- mean(data_4_7$language_likelihoodEFF)

data_4_7$language_likelihoodEFF.sum <- ifelse(data_4_7$language_likelihoodEFF == "1", 1-language_likelihoodEFF_mean, ifelse(data_4_7$language_likelihoodEFF == "0", 0-language_likelihoodEFF_mean, NA))

congruentEEF_mean <- mean(data_4_7$congruentEEF)

data_4_7$congruentEEF.sum <- ifelse(data_4_7$congruentEEF == "1", 1-congruentEEF_mean, ifelse(data_4_7$congruentEEF == "0", 0-congruentEEF_mean, NA))

#model

model.output1 = glmer(correct ~ language_likelihoodEFF*congruentEEF + (language_likelihoodEFF.sum * congruentEEF.sum||subjF) + (language_likelihoodEFF.sum * congruentEEF.sum||objectF), family = "binomial", data = data_4_7)

4.5)

#sum coding

data_partial$congruentEEF <- ifelse(data_partial$gaze_valid == data_partial$language_valid, 1, 0)

data_partial$congruent <- ifelse(data_partial$congruentEEF == 1, "Congruent", ifelse(data_partial$congruentEEF == 0, "Incongruent", NA))

gaze_likelihoodEFF_mean <- mean(data_partial$gaze_likelihoodEFF)

data_partial$gaze_likelihood.sum <- ifelse(data_partial$gaze_likelihoodEFF == "1", 1-gaze_likelihoodEFF_mean, ifelse(data_partial$gaze_likelihoodEFF == "0", 0-gaze_likelihoodEFF_mean, NA))

congruentEEF_mean <- mean(data_partial$congruentEEF)

data_partial$congruentEEF.sum <- ifelse(data_partial$congruentEEF == "1", 1-congruentEEF_mean, ifelse(data_partial$congruentEEF == "0", 0-congruentEEF_mean, NA))

#model

```
model.output1 = glmer(correct ~ gaze_likelihood.sum*congruentEEF.sum + (gaze_likelihood.sum*congruentEEF.sum||subjF) + (0+gaze_likelihood.sum|objectF), family = "binomial", data = data_partial)
```

4.6)

#sum coding

```
data_partial$congruentEEF <- ifelse(data_partial$gaze_valid == data_partial$language_valid, 1, 0 )

data_partial$congruent <- ifelse(data_partial$congruentEEF == 1, "Congruent", ifelse(data_partial$congruentEEF == 0, "Incongruent", NA))

gaze_likelihoodEFF_mean <- mean(data_partial$gaze_likelihoodEFF)

data_partial$gaze_likelihood.sum <- ifelse(data_partial$gaze_likelihoodEFF == "1", 1-gaze_likelihoodEFF_mean, ifelse(data_partial$gaze_likelihoodEFF == "0", 0-gaze_likelihoodEFF_mean, NA))

congruentEEF_mean <- mean(data_partial$congruentEEF)

data_partial$congruentEEF.sum <- ifelse(data_partial$congruentEEF == "1", 1-congruentEEF_mean, ifelse(data_partial$congruentEEF == "0", 0-congruentEEF_mean, NA))
```

#model

```
model.output1 = glmer(correct ~ gaze_likelihood.sum*congruentEEF.sum + (gaze_likelihood.sum*congruentEEF.sum||subjF) + (congruentEEF.sum||objectF), family = "binomial", data = data_partial)
```

4.7)

#sum coding

```
data$gaze_likelihoodEFF <- ifelse(data$session == "2", 1, ifelse(data$session == "3", 0, 9))
```

data_2_3 <- data[data$gaze_likelihoodEFF != "9",]

data_2_3$congruentEEF <- ifelse(data_2_3$gaze_valid == data_2_3$language_valid, 1, 0)

data_2_3$congruent <- ifelse(data_2_3$congruentEEF == 1, "Congruent", ifelse(data_2_3$congruentEEF == 0, "Incongruent", NA))

gaze_likelihoodEFF_mean <- mean(data_2_3$gaze_likelihoodEFF)

data_2_3$gaze_likelihood.sum <- ifelse(data_2_3$gaze_likelihoodEFF == "1", 1-gaze_likelihoodEFF_mean, ifelse(data_2_3$gaze_likelihoodEFF == "0", 0-gaze_likelihoodEFF_mean, NA))

congruentEEF_mean <- mean(data_2_3$congruentEEF)

data_2_3$congruentEEF.sum <- ifelse(data_2_3$congruentEEF == "1", 1-congruentEEF_mean, ifelse(data_2_3$congruentEEF == "0", 0-congruentEEF_mean, NA))

#model

model.output1 = glmer(first_direction_correct ~ gaze_likelihood.sum*congruentEEF.sum + (gaze_likelihood.sum* congruentEEF.sum||subjF) + (0+congruentEEF.sum||objectF), family = "binomial", data = data_2_3)

4.8)

#sum coding

data$language_likelihoodEFF <- ifelse(data$session == "4", 1, ifelse(data$session == "7", 0, 9))

data_4_7 <- data[data$language_likelihoodEFF != "9",]

data_4_7$congruentEEF <- ifelse(data_4_7$gaze_valid == data_4_7$language_valid, 1, 0)

data_4_7$congruent <- ifelse(data_4_7$congruentEEF == 1, "Congruent", ifelse(data_4_7$congruentEEF == 0, "Incongruent", NA))

language_likelihoodEFF_mean <- mean(data_4_7$language_likelihoodEFF) data_4_7$language_likelihoodEFF.sum <- ifelse(data_4_7$language_likelihoodEFF == "1", 1-language_likelihoodEFF_mean, ifelse(data_4_7$language_likelihoodEFF == "0", 0-language_likelihoodEFF_mean, NA))

congruentEEF_mean <- mean(data_4_7$congruentEEF)

data_4_7$congruentEEF.sum <- ifelse(data_4_7$congruentEEF == "1", 1-congruentEEF_mean, ifelse(data_4_7$congruentEEF == "0", 0-congruentEEF_mean, NA))

#model

model.output1 = glmer(first_direction_correct ~ language_likelihoodEFF*congruentEEF + (language_likelihoodEFF.sum * congruentEEF.sum||subjF) + (language_likelihoodEFF.sum * congruentEEF.sum||objectF), family = "binomial", data = data_4_7)

4.9)

#sum coding

data2$gaze_likelihoodEFF <- ifelse(data2$session == "2", 1, ifelse(data2$session == "3", 0, 9))

data_2_3 <- data2[data2$gaze_likelihoodEFF != "9",]

data_2_3$congruentEEF <- ifelse(data_2_3$gaze_valid == data_2_3$language_valid, 1, 0)

data_2_3$congruent <- ifelse(data_2_3$congruentEEF == 1, "Congruent", ifelse(data_2_3$congruentEEF == 0, "Incongruent", NA))

gaze_likelihoodEFF_mean <- mean(data_2_3$gaze_likelihoodEFF)

data_2_3$gaze_likelihood.sum <- ifelse(data_2_3$gaze_likelihoodEFF == "1", 1-gaze_likelihoodEFF_mean, ifelse(data_2_3$gaze_likelihoodEFF == "0", 0-gaze_likelihoodEFF_mean, NA))

congruentEEF_mean <- mean(data_2_3$congruentEEF) data_2_3$congruentEEF.sum <- ifelse(data_2_3$congruentEEF == "1", 1-congruentEEF_mean, ifelse(data_2_3$congruentEEF == "0", 0-congruentEEF_mean, NA))

#model

model.output2 = lmer(first_latency ~ gaze_likelihood.sum*congruentEEF.sum + (gaze_likelihood.sum*congruentEEF.sum|subjF)+(gaze_likelihood.sum*congruentEEF.sum|objectF), data = data_2_3)

4.10)

#sum coding

data2$language_likelihoodEFF <- ifelse(data2$session == "4", 1, ifelse(data2$session == "7", 0, 9))

data_4_7 <- data2[data2$language_likelihoodEFF != "9",]

data_4_7$congruentEEF <- ifelse(data_4_7$gaze_valid == data_4_7$language_valid, 1, 0)

data_4_7$congruent <- ifelse(data_4_7$congruentEEF == 1, "Congruent", ifelse(data_4_7$congruentEEF == 0, "Incongruent", NA))

language_likelihoodEFF_mean <- mean(data_4_7$language_likelihoodEFF)

data_4_7$language_likelihoodEFF.sum <- ifelse(data_4_7$language_likelihoodEFF == "1", 1-language_likelihoodEFF_mean, ifelse(data_4_7$language_likelihoodEFF == "0", 0-language_likelihoodEFF_mean, NA))

congruentEEF_mean <- mean(data_4_7$congruentEEF)

data_4_7$congruentEEF.sum <- ifelse(data_4_7$congruentEEF == "1", 1-congruentEEF_mean, ifelse(data_4_7$congruentEEF == "0", 0-congruentEEF_mean, NA))

#model

model.output2 = lmer(first_latency ~ language_likelihoodEFF.sum*congruentEEF.sum + (language_likelihoodEFF.sum*congruentEEF.sum|subjF) + (language_likelihoodEFF.sum*congruentEEF.sum||objectF), data = data_4_7)

4.11)

#sum coding

```
data2$gaze_likelihoodEFF <- ifelse(data2$session == "2", 1, ifelse(data2$session == "3", 0, 9))

data_2_3 <- data2[data2$gaze_likelihoodEFF != "9",]

data_2_3$congruentEEF <- ifelse(data_2_3$gaze_valid == data_2_3$language_valid, 1, 0 )

data_2_3$congruent <- ifelse(data_2_3$congruentEEF == 1, "Congruent", ifelse(data_2_3$congruentEEF == 0, "Incongruent", NA))


gaze_likelihoodEFF_mean <- mean(data_2_3$gaze_likelihoodEFF)

data_2_3$gaze_likelihood.sum <- ifelse(data_2_3$gaze_likelihoodEFF == "1", 1-gaze_likelihoodEFF_mean, ifelse(data_2_3$gaze_likelihoodEFF == "0", 0-gaze_likelihoodEFF_mean, NA))


congruentEEF_mean <- mean(data_2_3$congruentEEF)

data_2_3$congruentEEF.sum <- ifelse(data_2_3$congruentEEF == "1", 1-congruentEEF_mean, ifelse(data_2_3$congruentEEF == "0", 0-congruentEEF_mean, NA))
```

#model

```
model.output2 = lmer(time_to_hit_target ~ gaze_likelihood.sum*congruentEEF.sum + (gaze_likelihood.sum*congruentEEF.sum|subjF)+(gaze_likelihood.sum*congruentEEF.sum|objectF), data = data_2_3)
```

4.12)

sum coding

```
data2$language_likelihoodEFF <- ifelse(data2$session == "4", 1, ifelse(data2$session == "7", 0, 9))

data_4_7 <- data2[data2$language_likelihoodEFF != "9",]

data_4_7$congruentEEF <- ifelse(data_4_7$gaze_valid == data_4_7$language_valid, 1, 0 )

data_4_7$congruent <- ifelse(data_4_7$congruentEEF == 1, "Congruent", ifelse(data_4_7$congruentEEF == 0, "Incongruent", NA))
```

```
language_likelihoodEFF_mean <- mean(data_4_7$language_likelihoodEFF)

data_4_7$language_likelihoodEFF.sum <- ifelse(data_4_7$language_likelihoodEFF
== "1", 1-language_likelihoodEFF_mean, ifelse(data_4_7$language_likelihoodEFF
== "0", 0-language_likelihoodEFF_mean, NA))


congruentEEF_mean <- mean(data_4_7$congruentEEF)

data_4_7$congruentEEF.sum <- ifelse(data_4_7$congruentEEF == "1", 1-
congruentEEF_mean, ifelse(data_4_7$congruentEEF == "0", 0-congruentEEF_mean,
NA))
```

#model

```
model.output2 = lmer(time_to_hit_target ~
language_likelihoodEFF.sum*congruentEEF.sum +
(language_likelihoodEFF.sum*congruentEEF.sum|subjF) +
(language_likelihoodEFF.sum*congruentEEF.sum||objectF), data = data_4_7)
```

4.13)

#sum coding

```
data2$gaze_likelihoodEFF <- ifelse(data2$session == "2", 1, ifelse(data2$session ==
"3", 0, 9))

data_2_3 <- data2[data2$gaze_likelihoodEFF != "9",]

data_2_3$congruentEEF <- ifelse(data_2_3$gaze_valid == data_2_3$language_valid,
1, 0 )

data_2_3$congruent <- ifelse(data_2_3$congruentEEF == 1, "Congruent",
ifelse(data_2_3$congruentEEF == 0, "Incongruent", NA))


gaze_likelihoodEFF_mean <- mean(data_2_3$gaze_likelihoodEFF)

data_2_3$gaze_likelihood.sum <- ifelse(data_2_3$gaze_likelihoodEFF == "1", 1-
gaze_likelihoodEFF_mean, ifelse(data_2_3$gaze_likelihoodEFF == "0", 0-
gaze_likelihoodEFF_mean, NA))


congruentEEF_mean <- mean(data_2_3$congruentEEF)
```

data_2_3$congruentEEF.sum <- ifelse(data_2_3$congruentEEF == "1", 1-congruentEEF_mean, ifelse(data_2_3$congruentEEF == "0", 0-congruentEEF_mean, NA))

#model

model.output2 = lmer(RT ~ gaze_likelihood.sum*congruentEEF.sum + (gaze_likelihood.sum*congruentEEF.sum|subjF) +(gaze_likelihood.sum*congruentEEF.sum||objectF), data = data_2_3)

4.14)

#sum coding

data2$language_likelihoodEFF <- ifelse(data2$session == "4", 0, ifelse(data2$session == "7", 1, 9))

data_4_7 <- data2[data2$language_likelihoodEFF != "9",]

data_4_7$congruentEEF <- ifelse(data_4_7$gaze_valid == data_4_7$language_valid, 1, 0)

data_4_7$congruent <- ifelse(data_4_7$congruentEEF == 1, "Congruent", ifelse(data_4_7$congruentEEF == 0, "Incongruent", NA))

language_likelihoodEFF_mean <- mean(data_4_7$language_likelihoodEFF)

data_4_7$language_likelihoodEFF.sum <- ifelse(data_4_7$language_likelihoodEFF == "1", 1-language_likelihoodEFF_mean, ifelse(data_4_7$language_likelihoodEFF == "0", 0-language_likelihoodEFF_mean, NA))

congruentEEF_mean <- mean(data_4_7$congruentEEF)

data_4_7$congruentEEF.sum <- ifelse(data_4_7$congruentEEF == "1", 1-congruentEEF_mean, ifelse(data_4_7$congruentEEF == "0", 0-congruentEEF_mean, NA))

#model

model.output2 = lmer(RT ~ language_likelihoodEFF.sum*congruentEEF.sum + (language_likelihoodEFF.sum*congruentEEF.sum|subjF) +(language_likelihoodEFF.sum*congruentEEF.sum||objectF), data = data_4_7)

CHAPTER 5

5.1)

#sum coding

```
meangaze <- mean(data_2_3$gaze_likelihoodEFF)

data_2_3$gaze_likelihood.sum <- data_2_3$gaze_likelihoodEFF-meangaze


meancong <- mean(data_2_3$congruentEEF)

data_2_3$congruent.sum <- data_2_3$congruentEEF-meancong
```

#model

```
model.output1 = glmer(correct ~ gaze_likelihood.sum*congruent.sum + (gaze_likelihood.sum + congruent.sum||subjF)+(gaze_likelihood.sum||objectF),family = "binomial", data = data_2_3)
```

5.2)

#sum coding

```
meanlang <- mean(data_4_7$language_likelihoodEFF)

data_4_7$lang_likelihood.sum <- data_4_7$language_likelihoodEFF-meanlang


meancong <- mean(data_4_7$congruentEEF)

data_4_7$congruent.sum <- data_4_7$congruentEEF-meancong
```

#model

```
model.output1 = glmer(correct ~ lang_likelihood.sum*congruent.sum + (lang_likelihood.sum + congruent.sum||subjF)+ (lang_likelihood.sum||objectF), family = "binomial", data = data_4_7)
```

5.3)

#sum coding

```
meangaze <- mean(data_2_3$gaze_likelihoodEFF)

data_2_3$gaze_likelihood.sum <- data_2_3$gaze_likelihoodEFF-meangaze


meancong <- mean(data_2_3$congruentEEF)

data_2_3$congruent.sum <- data_2_3$congruentEEF-meancong
```

#model

```
model.output1 = glmer(first_direction_correct ~ gaze_likelihood.sum*congruent.sum
  + (gaze_likelihood.sum + congruent.sum||subjF)+(gaze_likelihood.sum *
  congruent.sum||objectF),family = "binomial", data = data_2_3)
```

5.4)

#sum coding

```
meanlang <- mean(data_4_7$language_likelihoodEFF)

data_4_7$lang_likelihood.sum <- data_4_7$language_likelihoodEFF-meanlang


meancong <- mean(data_4_7$congruentEEF)

data_4_7$congruent.sum <- data_4_7$congruentEEF-meancong
```

#model

```
model.output1 = glmer(first_direction_correct ~ lang_likelihood.sum*congruent.sum
  + (lang_likelihood.sum + congruent.sum||subjF)+ (0+lang_likelihood.sum|objectF),
  family = "binomial", data = data_4_7)
```

5.5)

#sum coding

```
meangaze <- mean(data_2_3$gaze_likelihoodEFF)

data_2_3$gaze_likelihood.sum <- data_2_3$gaze_likelihoodEFF-meangaze


meancong <- mean(data_2_3$congruentEEF)

data_2_3$congruent.sum <- data_2_3$congruentEEF-meancong
```

#model

```
model.output2 = lmer(first_latency ~ gaze_likelihood.sum*congruent.sum +
(gaze_likelihood.sum*congruent.sum|subjF) +
(gaze_likelihood.sum*congruent.sum|objectF), data = data_2_3)

summary(model.output2)
```

5.6)

#sum coding

```
meanlang <- mean(data_4_7$language_likelihoodEFF)

data_4_7$lang_likelihood.sum <- data_4_7$language_likelihoodEFF-meanlang


meancong <- mean(data_4_7$congruentEEF)

data_4_7$congruent.sum <- data_4_7$congruentEEF-meancong
```

#model

```
model.output2 = lmer(first_latency ~ lang_likelihood.sum*congruent.sum +
(lang_likelihood.sum*congruent.sum|subjF) +
(lang_likelihood.sum*congruent.sum||objectF), data = data_4_7)
```

5.7)

#sum coding

```
meangaze <- mean(data_2_3$gaze_likelihoodEFF)

data_2_3$gaze_likelihood.sum <- data_2_3$gaze_likelihoodEFF-meangaze


meancong <- mean(data_2_3$congruentEEF)

data_2_3$congruent.sum <- data_2_3$congruentEEF-meancong
```

#model

```
model.output2 = lmer(time_to_hit_target ~ gaze_likelihood.sum*congruent.sum +
                    (gaze_likelihood.sum*congruent.sum|subjF) +
          (gaze_likelihood.sum*congruent.sum||objectF), data = data_2_3)
```

5.8)

#sum coding

```
meanlang <- mean(data_4_7$language_likelihoodEFF)

data_4_7$lang_likelihood.sum <- data_4_7$language_likelihoodEFF-meanlang


meancong <- mean(data_4_7$congruentEEF)

data_4_7$congruent.sum <- data_4_7$congruentEEF-meancong
```

#model

```
model.output2 = lmer(time_to_hit_target ~ lang_likelihood.sum*congruent.sum +
                    (lang_likelihood.sum*congruent.sum|subjF) +
          (lang_likelihood.sum*congruent.sum|objectF), data = data_4_7)
```

5.9)

#sum coding

```
meangaze <- mean(data_2_3$gaze_likelihoodEFF)

data_2_3$gaze_likelihood.sum <- data_2_3$gaze_likelihoodEFF-meangaze
```

meancong <- mean(data_2_3$congruentEEF)

data_2_3$congruent.sum <- data_2_3$congruentEEF-meancong

#model

model.output2 = lmer(RT ~ gaze_likelihood.sum*congruent.sum + (gaze_likelihood.sum*congruent.sum|subjF) + (gaze_likelihood.sum*congruent.sum|objectF), data = data_2_3)

5.10)

#sum coding

meanlang <- mean(data_4_7$language_likelihoodEFF)

data_4_7$lang_likelihood.sum <- data_4_7$language_likelihoodEFF-meanlang

meancong <- mean(data_4_7$congruentEEF)

data_4_7$congruent.sum <- data_4_7$congruentEEF-meancong

#model

model.output2 = lmer(RT ~ lang_likelihood.sum*congruent.sum + (lang_likelihood.sum*congruent.sum|subjF) + (lang_likelihood.sum*congruent.sum|objectF), data = data_4_7)

5.11)

#sum coding

meancue <- mean(data$cue_condition-1)

data$cue_condition.sum <-data$cue_condition-1-meancue

#model

model.output1 <- glmer(correct ~ cue_condition.sum + (cue_condition.sum|subjF) + (cue_condition.sum||objectF), family = binomial, data = data)

summary(model.output1)

5.12)

#sum coding

```
meanswitch <- mean(data5$switch_code)

data5$switch.sum <-data5$switch_code-meanswitch


meanref <- mean(data5$reference-1)

data5$reference.sum <- data5$reference -1- meanref
```

#model

```
model.output1 <- glmer(correct ~ reference.sum*switch.sum + (reference.sum+switch.sum||subjF) + (0+reference.sum||objectF), family = binomial, data = data5)
```

5.13)

#sum coding

```
meancue <- mean(data$cue_condition-1)

data$cue_condition.sum <-data$cue_condition-1-meancue
```

#model

```
model.output1 <- glmer(first_direction_correct ~ cue_condition.sum + (cue_condition.sum|subjF) + (cue_condition.sum||objectF), family = binomial, data = data)
```

5.14)

#sum coding

```
meanswitch <- mean(data5$switch_code)

data5$switch.sum <-data5$switch_code-meanswitch


meanref <- mean(data5$reference-1)

data5$reference.sum <- data5$reference -1- meanref
```

#model

```
model.output1 <- glmer(first_direction_correct ~ reference.sum*switch.sum + (reference.sum+switch.sum||subjF) + (1|objectF), family = binomial, data = data5)
```

5.15)

#sum coding

```
meancue <- mean(data4$cue_condition-1)

data4$cue_condition.sum <-data4$cue_condition-1-meancue
```

#model

```
model.output1 <- lmer(first_latency ~ cue_condition.sum + (cue_condition.sum|subjF) + (cue_condition.sum|objectF), data = data4)
```

5.16)

#sum coding

```
meanswitch <- mean(data5$switch_code)

data5$switch.sum <-data5$switch_code-meanswitch


meanref <- mean(data5$reference-1)

data5$reference.sum <- data5$reference -1- meanref
```

#model

```
model.output1 <- lmer(first_latency ~ reference.sum*switch.sum + (reference.sum*switch.sum|subjF) + (reference.sum*switch.sum|objectF), data = data5)
```

5.17)

#sum coding

```
meancue <- mean(data4$cue_condition-1)

data4$cue_condition.sum <-data4$cue_condition-1-meancue
```

#model

```
model.output1 <- lmer(time_to_hit_target ~ cue_condition.sum + (cue_condition.sum|subjF) + (cue_condition.sum|objectF), data = data4)
```

5.18)

#sum coding

```
meanswitch <- mean(data5$switch_code)

data5$switch.sum <-data5$switch_code-meanswitch


meanref <- mean(data5$reference-1)

data5$reference.sum <- data5$reference -1- meanref
```

#model

```
model.output1 <- lmer(time_to_hit_target ~ reference.sum*switch.sum + (reference.sum*switch.sum|subjF) + (reference.sum*switch.sum|objectF), data = data5)
```

5.19)

#sum coding

```
meancue <- mean(data4$cue_condition-1)

data4$cue_condition.sum <-data4$cue_condition-1-meancue
```

#model

```
model.output1 <- lmer(RT ~ cue_condition.sum + (cue_condition.sum|subjF) + (cue_condition.sum|objectF), data = data4)
```

5.20)

#sum coding

```
meanswitch <- mean(data5$switch_code)

data5$switch.sum <-data5$switch_code-meanswitch
```

meanref <- mean(data5$reference-1)

data5$reference.sum <- data5$reference -1- meanref

#model

model.output1 <- lmer(RT ~ reference.sum*switch.sum + (reference.sum*switch.sum|subjF) + (reference.sum*switch.sum|objectF), data = data5)

Appendix 3: Victoria Sponge recipe

Recipe Procedure

1) Gather the large bowl and small bowl
2) Gather 50g of butter
3) Put butter in the small bowl
4) Put the small bowl in the microwave for 15 seconds on high
5) Add butter to large bowl
6) Gather 100g of sugar
7) Add sugar to large bowl
8) Get whisk
9) Whisk butter and sugar until light and fluffy
10) Gather 2 eggs
11) Add the eggs to large bowl and begin to whisk
12) Gather the sieve **WHILST** continuing to whisk
13) Gather 100g of flour **WHILST** continuing to whisk
14) Sieve the flour a little bit at a time **WHILST** continuing to whisk
15) Gather milk **WHILST** continuing to whisk
16) Add a small amount of milk to bowl **WHILST** continuing to whisk
17) Once everything is evenly mixed, stop whisking.
18) Gather the square tin
19) Add mixture to square tin
20) Spread the mixture evenly
21) Put the square tin in oven

THE RECIPE IS NOW COMPLETE

Appendix 4: Chef guidelines

You are the chef.

Your fellow participant is the **gatherer**.

As chef, you will ultimately be in charge of preparing the recipe. The gatherer is present to assist you.

You are expected to mix and prepare the ingredients, as outlined in the recipe. You are not expected to gather together the ingredients/items/appliances that you need. However, as you are in charge, you are expected to delegate this responsibility to the gatherer.

You will be given a recipe to follow for Victoria Sponge. You cannot show the recipe to the gatherer, but you are free to share any of the information it contains. You will be familiarised with the kitchen before the procedure begins, without the gatherer.

If there is any stage of the mixing/preparation that you feel could be carried out more easily with two people, you can get the gatherer to assist you. However, as you are in charge you will be responsible for explicitly guiding the gatherer through these stages.

If you have any questions about your responsibilities as the chef, feel free to ask the experimenter at this point.

Once you and the gatherer are happy that you understand your individual roles, the experimenter will give you the recipe and you may begin the procedure.

Appendix 5: Gatherer guidelines

You are the gatherer.

Your fellow participant is the **chef**.

The chef is ultimately in charge of preparing the recipe. You are present to assist the chef.

You will not be expected to make any decisions regarding the preparation. Your primary responsibility will be to collect ingredients/items/appliances that the chef requires to prepare the recipe.

The chef will be given a recipe for Victoria Sponge. You must not read this. Instead, you will get information directly from the chef.

If the chef asks you to assist in the mixing/preparation of the ingredients you should do so. However, you must remember to closely follow the instructions of the chef.

If you have any questions about your responsibilities as the gatherer, feel free to ask the experimenter at this point.

Once you and the chef are happy that you understand your individual roles, the experimenter will give the recipe to the chef and you both may begin the procedure.

Appendix 6: Autism Quotient Questionnaire

This questionnaire was based on the Autism-spectrum Quotient (AQ) questionnaire from the appendix of Baron-Cohen et al. (2001). The differences between this Questionnaire and the questionnaire used by Baron-Cohen et al (2001) are all superficial – The exact same question statements are used. Participants were scored out of fifty for autistic traits, with each question getting a score of 1 or 0. Responses of "Definitely agree" and "Slightly agree" scored 1 for questions 1, 2, 4, 5, 6, 7, 9, 12, 13, 16, 18, 19, 20, 21, 22, 23, 26, 33, 35, 39, 41, 42, 43, 45 and 46. Responses of "Definitely disagree" and "Slightly disagree" scored 1 for questions 3, 8, 10, 11, 14, 15, 17, 24, 25, 27, 28, 29, 30, 31, 32, 34, 36, 37, 38, 40, 44, 47, 48, 49 and 50. The higher the score, the higher the participant's autistic traits.

The five pages of the Autism Quotient Questionnaire are shown below:

	Definitely agree	Slightly agree	Slightly disagree	Definitely disagree
1. I prefer to do things with others rather than on my own.	☐	☐	☐	☐
2. I prefer to do things the same way over and over again.	☐	☐	☐	☐
3. If I try to imagine something, I find it very easy to create a picture in my mind.	☐	☐	☐	☐
4. I frequently get so strongly absorbed in one thing that I lose sight of other things.	☐	☐	☐	☐
5. I often notice small sounds when others do not.	☐	☐	☐	☐
6. I usually notice car number plates or similar strings of information.	☐	☐	☐	☐
7. Other people frequently tell me that what I've said is impolite, even though I think it is polite.	☐	☐	☐	☐
8. When I'm reading a story, I can easily imagine what the characters might look like	☐	☐	☐	☐
9. I am fascinated by dates.	☐	☐	☐	☐
10. In a social group, I can easily keep track of several different people's conversations.	☐	☐	☐	☐

	Definitely agree	Slightly agree	Slightly disagree	Definitely disagree
11. I find social situations easy.	☐	☐	☐	☐
12. I tend to notice details that others do not.	☐	☐	☐	☐
13. I would rather go to a library than a party.	☐	☐	☐	☐
14. I find making up stories easy.	☐	☐	☐	☐
15. I find myself drawn more strongly to people than to things.	☐	☐	☐	☐
16. I tend to have very strong interests, which I get upset about if I can't pursue.	☐	☐	☐	☐
17. I enjoy social chit-chat.	☐	☐	☐	☐
18. When I talk, it isn't always easy for others to get a word in edgeways.	☐	☐	☐	☐
19. I am fascinated by numbers.	☐	☐	☐	☐
20. When I'm reading a story, I find it difficult to work out the characters' intentions.	☐	☐	☐	☐

	Definitely agree	Slightly agree	Slightly disagree	Definitely disagree
21. I don't particularly enjoy reading fiction.	☐	☐	☐	☐
22. I find it hard to make new friends.	☐	☐	☐	☐
23. I notice patterns in things all the time.	☐	☐	☐	☐
24. I would rather go to the theatre than a museum.	☐	☐	☐	☐
25. It does not upset me if my daily routine is disturbed.	☐	☐	☐	☐
26. I frequently find that I don't know how to keep a conversation going.	☐	☐	☐	☐
27. I find it easy to "read between the lines" when someone is talking to me.	☐	☐	☐	☐
28. I usually concentrate more on the whole picture, rather than the small details.	☐	☐	☐	☐
29. I am not very good at remembering phone numbers.	☐	☐	☐	☐
30. I don't usually notice small changes in a situation, or a person's appearance.	☐	☐	☐	☐

	Definitely agree	Slightly agree	Slightly disagree	Definitely disagree
31. I know how to tell if someone listening to me is getting bored.	☐	☐	☐	☐
32. I find it easy to do more than one thing at once.	☐	☐	☐	☐
33. When I talk on the phone, I'm not sure when it's my turn to speak.	☐	☐	☐	☐
34. I enjoy doing things spontaneously.	☐	☐	☐	☐
35. I am often the last to understand the point of a joke.	☐	☐	☐	☐
36. I find it easy to work out what someone is thinking or feeling just by looking at their face.	☐	☐	☐	☐
37. If there is an interruption, I can switch back to what I was doing very quickly.	☐	☐	☐	☐
38. I am good at social chit-chat.	☐	☐	☐	☐
39. People often tell me that I keep going on and on about the same thing.	☐	☐	☐	☐
40. When I was young, I used to enjoy playing games involving pretending with other children.	☐	☐	☐	☐

	Definitely agree	Slightly agree	Slightly disagree	Definitely disagree
41. I like to collect information about categories of things (e.g. types of car, types of bird, types of train, types of plant, etc.).	☐	☐	☐	☐
42. I find it difficult to imagine what it would be like to be someone else.	☐	☐	☐	☐
43. I like to plan any activities I participate in carefully.	☐	☐	☐	☐
44. I enjoy social occasions.	☐	☐	☐	☐
45. I find it difficult to work out people's intentions.	☐	☐	☐	☐
46. New situations make me anxious.	☐	☐	☐	☐
47. I enjoy meeting new people.	☐	☐	☐	☐
48. I am a good diplomat.	☐	☐	☐	☐
49. I am not very good at remembering people's date of birth.	☐	☐	☐	☐
50. I find it very easy to play games with children that involve pretending.	☐	☐	☐	☐

Appendix 7 – Picture Stimuli for Studies 3, 4 and 5


Green air freshener


Red air freshener


Blue boomerang


Green boomerang


Blue brush


Yellow brush


Red calculator


Yellow calculator


Orange candle holder


Yellow candle holder


Brown cat


White cat


Green clothes-hanger


Yellow clothes-hanger


Blue cowboy boot


Red cowboy boot


Black diary


Brown diary


Blue fork


Orange fork


Green frisbee


Yellow frisbee


Blue hat


Pink hat


Orange highlighter


Yellow highlighter


Beige jug


Brown jug


Blue knife


Orange knife


Blue lantern


Yellow lantern


Blue mat


Green mat


Blue pencil holder


Pink pencil holder


Green pencil sharpener


Pink pencil sharpener


Pink purse


White purse


Blue bow


White bow


Blue snowman


Pink snowman


Blue spoon


Orange spoon


Blue watch


Red watch


Black spray can


Pink spray can


Black sunglasses


White sunglasses


Blue toothpick holder


Green toothpick holder


Blue washing line


Red washing line


Green wine glass


Red wine glass


Black wool


Brown wool